HF5501
.R8

Columbia University

Contributions to Education

Teachers College Series

CMSC
Ward Edwards Library
Warrensburg, Missouri

Ex

DATE DUE

A DIAGNOSTIC TEST OF APTITUDE FOR CLERICAL OFFICE WORK

Based on an Analysis of Clerical Operations

BY

ALLEN MEAD RUGGLES, PH.D.

TEACHERS COLLEGE, COLUMBIA UNIVERSITY
CONTRIBUTIONS TO EDUCATION, NO. 148

PROPERTY OF TEACHERS COLLEGE PUBLICATIONS

Published by
Teachers College, Columbia University
New York City
1924

Copyright, 1924 by Teachers College, Columbia University

PROPERTY OF
TEACHERS COLLEGE
MANHATTANVILLE

HF 5501
.R8

ACKNOWLEDGMENTS

This study, in its present form, would not have been possible except for the many courtesies extended to the writer by the Wisconsin Civil Service Commission, with whom he was connected in the capacity of Service Examiner. A considerable portion of the material presented represents the results of projects undertaken for the Commission.

The opportunity of trying out the tests on different types of clerical workers was provided by several large business and industrial concerns whose officials cooperated in a most helpful way.

To Dr. Edward L. Thorndike the writer is indebted for inspiration and suggestions leading to the making of the study, and to the following for help and suggestions during the progress of the study: Dr. Henry A. Ruger, Dr. Herbert A. Toops, Dr. V. A. C. Henmon, Dr. L. L. Thurstone and Dr. Arthur I. Gates.

A. M. R.

76878

CONTENTS

Page

INTRODUCTION vii

Chap.

I. THE ANALYSIS OF CLERICAL DUTIES 1

Many Titles for Clerical Positions.
Analyzing Clerical Work into General Duties.
A Clerical Performance Scale.
How the Scale is Used.
Standardizing the Grade of Duties.
The Field of Clerical Duties.
Further Analysis and Selection of Samplings for the Tests.

II. THE SIX TESTS AND THEIR ADMINISTRATION 17

Descriptions of the Tests.
Reaction rather than Knowledge Tests.
Method of Administering the Tests.
The Preliminary Tryout.

III. THE CRITERIA 26

Difficulties in Securing Adequate Criteria.
A Work Analysis for Each Group.
Criteria Ratings.
The Validity and Reliability of Criteria.
Factors Tending to Affect the Reliability of Ratings.
The Objective or Pure Output Criterion.
Bonus Records as Criteria.
Validity of a Given Weight for Errors in Determining Bonus.
Variation in the Judgment of Experienced Examiners.
Criteria Statistically Considered.
The Criteria of This Study.
Application to Civil Service Procedure.

IV. THE SCORING AND WEIGHTING PROBLEM 43

SECTION I

The Scoring, or Weighting for Errors

The Problems in a Clerical Test.
Attacking the Problem Statistically.
A study of Weighting Curves.
A testing for Consistency in the Curves.

SECTION 2
The Problem of Selecting and Weighting Tests for a Battery
Resulting Regression Equations.

V. RESULTING CORRELATIONS AND INTERPRETATIONS . . . 57
Methods and Resulting Data.
The Possible Use of the Tests in the Selection of General Clerks.
Consideration of the Tests for Prognosis of Specialized Clerical Work.
Aptitude or Trade Tests.
Suggestions for Improving the Tests.
Standards for Judging Test Performance.

VI. SUMMARY 77

APPENDIX 81

INTRODUCTION

This study was undertaken with the hope of devising and standardizing a set of tests that might be used in the selection of clerical office workers of the lower grades. It was the idea that such a set might later be supplemented by other tests based on the same underlying principles but covering the higher ranges of clerical work. Further it was intended from the first to give the tests of this study in connection with a general intelligence examination.

The study was restricted to so-called sampling tests, made up of samples of actual clerical duties, and devised only after a careful and thorough analysis of the clerical field. The aim was to have the resulting tests actually cover a cross-section of clerical work. This study of clerical work developed into quite large proportions in itself, much larger than necessary for the tests of this particular study. It is hoped, however, that the results of the analysis of clerical duties may serve as the basis for further work along this line.

The development of clerical testing is in its first stages. In the spirit of suggestion and helpfulness, therefore, this study is presented, with the hope that others may find in it bits of useful and suggestive material, and may be led to further investigate the problems which here are hardly more than opened up and which demand thorough investigation.

CHAPTER I

THE ANALYSIS OF CLERICAL WORK

The tests with which this study is concerned being of the type known as "sampling tests," it was quite necessary that a thorough analysis be made of the work of the positions for which the tests were planned. Although this group of tests covers only part of the general field of clerical office work, yet the underlying principles upon which the tests are built are applicable to the entire range of clerical performance. As a result of these considerations the clerical field as a whole was at first considered and then narrowed down to the section for which the tests of the study were especially devised. In this discussion the same order of procedure is followed.

It is not difficult to distinguish between the work of an office clerk and that of a teacher, a policeman, a machinist, or any of the numerous other occupations listed in a census report or similar publication. And yet, when we attempt to define clerical office work, we are impressed with the uncertainty of its limitations. Within the office itself it is far from easy to say just which employes are really doing clerical work and which are not. An accountant, a statistician, or an executive secretary may be performing almost the same type of work as other employes who are called clerks. There is also the question of what is to be done in this classification with the stenographer-clerk, the typist, and the various calculating and time-saving-machine operators. In many cases employes belonging to these groups perform duties which, done by another employe without the use of any machine, would unquestionably be called clerical in nature. In these cases skill in the use of a machine is added to the clerical ability, with many of the mental reactions common to clerical work still remaining.

In general the clerical work of an office includes most of the work done in connection with the receipt, handling, examination, recording, care, acknowledgment, and treatment of reports, correspondence and requests, together with all papers and records pertaining to the activities and functions of the business with which the office is connected. Included under the term "clerk" are widely varying

types and grades of ability. One office clerk is not necessarily like another in type, any more than in the field of salesmanship a beginning saleswoman in a five-and-ten-cent store is like a high grade insurance or bond salesman. There are office clerks who are only eighth grade graduates; while on the other hand there are those who have graduated from college and who have had specialized training besides. There are clerks receiving salaries of $40 a month and there are those whose monthly pay envelope contains $500. In many instances a lack of proper classification of employes, and a looseness in the use of titles is a factor in the confusion so often found in this field.

MANY TITLES FOR CLERICAL POSITIONS

An examination of the clerical titles appearing on the payrolls of various business concerns, as well as on the payrolls of the public service, is sufficient to convince one of the need for a more thoroughgoing analysis of clerical work. The following are suggestive of the many titles assigned to clerks:

Billing clerk	Warrant clerk
Accounting clerk	Purchasing clerk
Sales clerk	Statistical clerk
Record clerk	Filing clerk
Order clerk	Vault clerk
Auditing clerk	Secretarial clerk
Payroll clerk	Confidential clerk
Mailing clerk	Complaint clerk
Subscription clerk	Certification clerk
Dividend clerk	Corporation clerk
Mail-order clerk	Employment clerk
Checking clerk	Freight clerk
Stock clerk	Traffic clerk
Clerk-typist	Clerk-stenographer
Examining clerk	Clerk-bookkeeper

Such a list might easily be extended to many times the length of that given. It will serve, however, to suggest the problem that confronts the worker in this field. It is seen at once that in many cases these titles simply refer to the type of material handled and are significant only as they fit into the organization of an individual office.

A number of questions immediately arise: Are these positions such that a person who is successful in one may with equal success

fill a number of others? If so, then what are the common elements present in such positions? Or, are these positions really quite different? Nothing in the titles at least furnishes an answer to these questions. Further investigation is necessary.

With large business concerns, where many clerks are employed, job specifications are generally used by the employment department in connection with the hiring of clerks for each type of work. In this connection, however, each type of position such as billing clerk, order clerk, etc., is usually considered by itself. All too infrequently are comparisons between one position and another made with any degree of exactitude, even for the purpose of securing equity in comparative salaries. In the job specifications are set forth the duties of the clerks on the particular job in question. These of course vary in the amount of detail given. Typical of a shorter one is the following:

> The order clerk checks the orders as they come from the receiving room against the slips turned in by the salesmen, totals orders for the day, and makes out a report sheet on the daily output of the different departments.

Such descriptions are usually in terms of the special phraseology of the business in question, and such statements, in terms of the procedure of the individual office, are often not very illuminating to an outsider.

In the field of public service a thoroughgoing effort has been made during the past few years to analyze occupations, and to state the associated duties in terms most ready of comparison, this being done as a basis for salary standardization and promotions, and for examination purposes. It was on the basis of such an analysis of clerical work that the sampling tests[1] set forth in this study were constructed. An attempt was made to go beyond the usual job analysis[2] statement of duties, in order to secure more generalized descriptions which might serve as standards, and thus might be applied to clerical positions in any office.

ANALYZING CLERICAL WORK INTO GENERAL DUTIES

The basis for this analysis was a detailed statement of duties from practically every clerk in the state service of Wisconsin. These statements were carefully examined, and a list of typical

[1] See Appendix, page 84, (1).
[2] See Appendix, page 85, (2).

duties selected. Naturally each clerk's duties were stated in terms of the work of the particular position held. Those duties seeming to be practically the same, even though stated in different phraseology, were brought together. When such a group was found to be homogeneous, so far as the fundamental operations and difficulties of the duties were concerned, a general statement for such duties was formulated. For example, in a number of offices were found clerks among whose other duties was that of recording the receipt of various papers, documents, or even fees. In one office the paper might be an income tax report, in another a school report, in another a birth certificate, and so on. The terminology in the different cases might vary considerably, yet the operations performed by the clerks be essentially the same. The clerk who could read the name on an income tax report, and check its receipt against another list, could also do the same thing for a school report. The description of such duties was therefore reduced to a statement such as the following:

> Checks by mark or date opposite appropriate item on list, the receipt of fees, reports, and other papers.

Such a statement holds for many duties which in different offices might be expressed in varying phraseology, but which are practically the same, so far as the fundamental operations are concerned. This process was carried out for all of the common clerical duties, the result being a list of about seventy-five so-called typical duties, stated in terms applicable to any office. This list later served as the basis for the study of the fundamental operations[1] involved in clerical work. At this stage, however, it was developed into a standard for properly classifying clerical positions as to the grade or degree of difficulty of the work. This was done as a first step toward salary determination, and as suggestive in the formulation of examinations for applicants for clerical positions in the state service.

For convenience in handling these some seventy-five statements of typical duties, they were arranged in tabular form. (The idea of this form or "Standard Classification" may be gotten from the chart on the opposite page. This was taken from the "Office Service Bulletin" issued by the Wisconsin Civil Service Commission). These typical duties were grouped in horizontal rows ac-

[1] See Appendix, page 85, (4).

STANDARD CLASSIFICATION OF CLERICAL DUTIES

(Tentative)

	SUB-CLERK	ASSISTANT CLERK	JUNIOR CLERK	SENIOR CLERK
Duties in connection with and related to	Duties: Mechanical in nature including the simplest kind of office work. Compensation: $[illegible] to $[illegible] per month.	Duties: Of a simple clerical nature. Accurate transference of data or other material, requiring no special skill, is the important factor in this work. Compensation: $[illegible] to $[illegible] per month.	Duties: Clerical duties calling for ability to analyze and to work with rather complicated relationships. The method used in many cases follows no definite rule. A knowledge of departmental procedure and policy is involved in many of the duties of this grade. Compensation: $[illegible] to $[illegible] per month.	Duties: Clerical duties which call for a comprehensive knowledge of departmental procedure and policy, and many of which require the exercise of initiative and independent judgment. Compensation: $[illegible] and up.
Correspondence	Delivers departmental messages. Stamps mail with date of receipt. Seals and stamps mail. Addresses envelopes. Distributes mail. Mails pamphlets, reports, etc. Answers 'phone calls. Gets printed matter and other material ready to go out, wrapping and tying. Separates correspondence from reports and other matter.	Inspects all addressed mail before it leaves the office. Sends out set form requests, acknowledgments, etc. Has charge of requests for printed matter. Answers requests for fixed information, filling in set forms. Answers common requests for fixed information, using form paragraphs. Opens and sorts mail.	Examines for errors of spelling, omitted words, and English, letters sent out from the department. Answers requests for frequently asked information requiring variable responses made by the selection and use of form paragraphs. Writes letters after being told main items or points to be incorporated. Answers requests for information requiring simple but variable responses, sending personal replies. Briefs correspondence.	Answers requests for explanation of rules, etc., sending personal replies. Advises with people who come to office for general information. Prepares subject matter for department advertising. Answers requests for information requiring knowledge of application of laws governing departmental procedure. Answers unusual letters in which a good amount of tact and judgment is required.
Filing and Indexing	Cuts and pastes clippings. Numbers cards, correspondence, reports, etc., serially.	Locates towns, cities, etc., geographically and writes classification on same. Files cards, correspondence, reports, etc., alphabetically, numerically, and chronologically, or geographically without further subdivision. Searches record files for desired information, following a simple scheme of classification. Makes out index cards for classified papers, records, etc., including the recording of items from several sources.	Files cards, correspondence, reports, etc., according to subjects using a decimal or special scheme of filing. Searches record files for desired information, requiring a thorough knowledge of cross references. Briefs reports, etc. Classifies reports, clippings, books and such material according to subjects, using cross references. Classifies material according to subjects and according to use to be made of material.	Devises or revises index and filing scheme to meet peculiar needs. Classifies material according to subject or subjects, calling for an extensive use of technical terms and a comprehensive knowledge of departmental procedure.
Statistical Work		Operates adding machine. Performs the simple mathematical operations in connection with tabulations. Operates slide rule to secure unit costs, percentages, etc. Selects a number of items from a larger number, locating them under appropriate headings on tabulation sheet.	Classifies or tabulates data appearing in groups of related items. Drafts curve from simple data. Classifies data of such complexity that in tabulating and in watching for inconsistencies, cumulative attention and memory play a prominent part.	Tabulates complex data. Studies data for significant indications. Writes up interpretation of tables.
Recording and Accounting		Copies written matter in full by hand. Copies lists of names in full onto cards, one name per card. Copies lists of names in full and in regular order onto a record book or other blank form. Selects items and copies in regular order from lists onto record book or blank form. Copies complete list of items into record book each item under appropriate headings.	Copies selected items from list into record book under appropriate headings. Prepares mailing lists, lists for publication, etc. Summarizes items appearing under common headings and transfers result to appropriate place in record book or other set form. Summarizes items appearing under different headings and transfers result to appropriate place in record book or other set form. Accumulates items from a complexity of data, summarizes same under two or more headings and transfers to regular form.	Checks vouchers for mathematical accuracy, correctness of classification, etc. Receives, checks and files incoming reports. Posts to detailed and general ledgers. Prepares estimates and figures for financial statements. Keeps detailed books of records, such as voucher registers, cash disbursement records, sales records, etc.
Examining and Certifying	Stamps licenses and other papers, already sorted	Checks by mark or date opposite appropriate name on list, the receipt of fees, reports and other papers. Makes out requisitions on regular form, files copy, passing original on to department head for signature. Keeps track of departmental office supplies, making out lists of those needed and presents requisition for same to department head for his signature. Is responsible for keeping up the supply	Examines information form noting whether all required questions are answered, and whether the right TYPE of answer is given. Issues permits, licenses, etc., upon receipt of form statement properly witnessed or requests from proper classes of people. Receives fees stipulated by law, gives receipt, keeps record of fees received and turns over currency at frequent intervals. Examines condensed reports or filing cards for discrepancies shown by relationships of various items.	Examines for significant variations and inconsistencies, reports of many items, some of which may be omitted, checking part of the items against the same items found in one of several other reports, tracing down any error. Judges of right of parties to licenses, permits, etc., after analyzing report or application from the standpoint of legality, relationship of items playing a large part. A knowledge of the affecting laws and of the accounting system of the parties applying is required.

cording to the general nature of the work: as "filing and indexing," "statistical work," "recording and accounting," "correspondence," and "examining and certifying." This horizontal grouping was purely arbitrary and for the sake of convenience in locating the items. In vertical columns, these duties were divided according to degree of difficulty—graded under headings of "sub-clerical," "assistant clerical," "junior clerical," and "senior clerical." Thus instead of license clerks, certification clerks, auditing clerks etc., each clerk's *duties* were located on this form, and on that basis a classification or title given which made the work of the clerk understandable to anyone else in the service. The duties of clerks were expressed in comparable terms. This then might serve as a basis for direct comparison and salary equity. It is, however, the generalization of the duties, and their systematic and comparable arrangement, which is here set forth for emphasis.

As an example of these general statements of typical duties, the following are taken from different parts of the clerical chart.

Sends out set form requests, acknowledgements, etc.

Answers requests for fixed information, filling in set forms.

Makes out index cards for classified papers, records, etc.

Classifies material according to subject, and according to use to be made of such material.

Selects a number of items from a larger number, locating them under appropriate heading on tabulation sheet.

Examines information forms, noting whether all required questions are answered, and whether right type of answer is given.

Tabulates complex data.

At this stage of the procedure it was found that although the general definition or statement of a duty met a real need, yet this same value of generality had its drawbacks. For instance, in the last of the above statements of typical duties, the question may fairly be asked: "What determines when data are complex?" It makes little difference whether the data have to do with agriculture, immigration, or taxation; but it is quite important that it is thoroughly understood what is meant by the adjective "complex." Here seemed to be the place where a very definitely illustrated duty from a specialized field might have a positive value. This suggestion was carried out, and resulted in a series of illustrations in the form of a clerical performance scale, which scale is discussed in the following pages.

A CLERICAL PERFORMANCE SCALE

In the clerical performance scale,[1] each typical duty is represented by a card 9½ x 14 inches. On each card is an exhibit—a sample of the work involved in that particular duty. These cards are arranged in the order of the relative difficulty of the work represented. For convenience of reference the cards are divided into the several groups, already mentioned in the preceding section. For example, the cards illustrating duties connected with filing and indexing are all in one group, and arranged from the simplest to the most difficult. On each card is indicated, in the upper left-hand corner, the group to which it belongs, and its place in that group: as F–4—fourth card in the filing and index group. In the upper right-hand corner is placed the grade to which the particular duty has been assigned as a basis for classification.

Making up the exhibit on each of these scale cards may be found about the same division of information. In general it is as follows:

a. Statement of duty in terms applicable to any office.
(Same as those in clerical chart of typical duties.)
b. The primary form, which is the start of the operation—a report, record or request.
c. The resulting form or completed record.
d. An explanation of the operations involved.

The charts on pages 8 and 10 give an idea of the makeup of the cards used in this scale.

In the correspondence group is a series of letters of the type that an employe might be called upon to answer. They range from the simplest to the very difficult. In the simplest type the clerk has only to fill in a date on a form card; while the more difficult require an intimate knowledge of the work of the office as a whole, skill and tact in answering the letter, and ability to use good English. On each card is a sample letter of request, and the answer. Under directions, are to be found indications of the necessary information that must be had, and the degree of care that must be exercised in framing the answer.

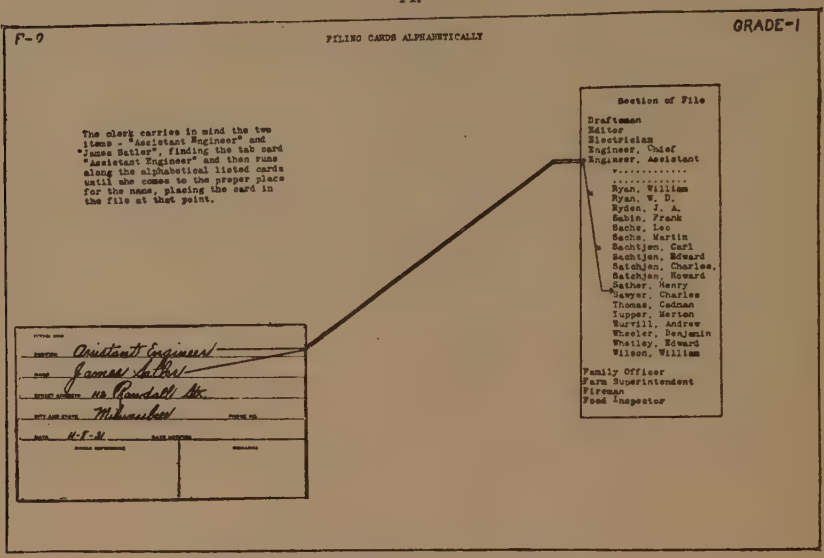
The clerical performance scale thus has all the advantages of the classified list of typical, generalized duties, together with that of concrete illustration. In other words, it is not only general in application, but it is concrete in definition. Each piece of work performed by a clerk, no matter in what type of office, may be

[1] See Appendix, page 85, (3).

CHART I

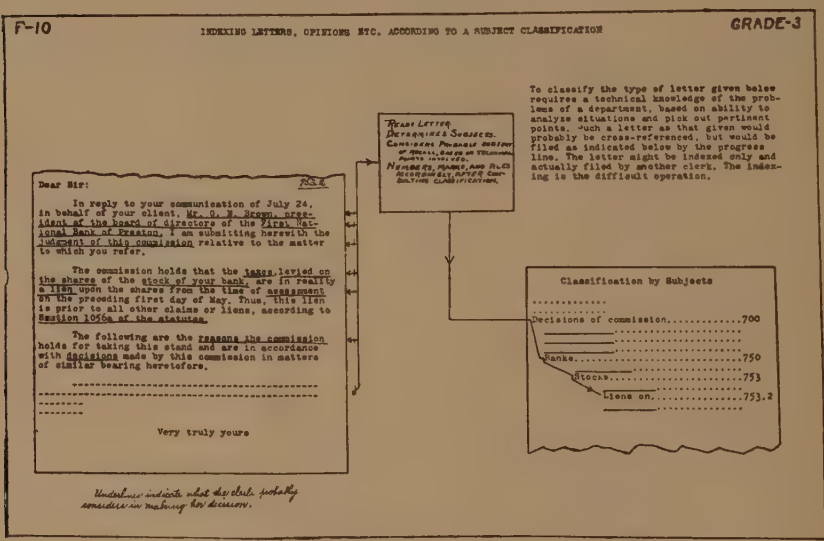
Illustration from Clerical Performance Scale[1]

A.



Second Card from Filing and Indexing Group

B.


F-10
INDEXING LETTERS, OPINIONS ETC. ACCORDING TO A SUBJECT CLASSIFICATION
GRADE-3
To classify the type of letter given below requires a technical knowledge of the problems of a department, based on ability to analyze situations and pick out pertinent points. Such a letter as that given would probably be cross-referenced, but would be filed as indicated below by the progress line. The letter might be indexed only and actually filed by another clerk. The indexing is the difficult operation.
Dear Sir:
In reply to your communication of July 24, in behalf of your client, Mr. C. M. Brown, president of the board of directors of the First National Bank of Preston, I am submitting herewith the judgment of this commission relative to the matter to which you refer.
The commission holds that the taxes levied on the shares of the stock of your bank are in reality a lien upon the shares from the time of assessment on the preceding first day of May. Thus, this lien is prior to all other claims or liens, according to Section 1056a of the statutes.
The following are the reasons the commission holds for taking this stand and are in accordance with decisions made by this commission in matters of similar bearing heretofore.
Very truly yours
Underlines indicate what the clerk probably considers in making her decision.
Classification by Subjects
Decisions of commission.....700
Banks.....750
Stocks.....753
Liens on.....753.2

Tenth Card from Filing and Indexing Group

[1] Ruggles, Allen M., "Classifying and Grading Individual Employments." Paper read at Fifteenth Annual Assembly of Civil Service Commissioners, San Francisco, June 6–9, 1922.

classified on the basis of this scale. It may be used in comparing the work of a clerk in one office with that of those in other offices, and in establishing a system of salary ranges based upon the principle of equal pay for equal work. That, however, which is of especial interest in connection with the clerical tests of this study is the opportunity which such a scale provides for studying and analyzing different types of work.

HOW THE SCALE IS USED

Let the scale be applied to a particular position which it is difficult to locate in a classification: The position involves a number of different duties, belonging, perhaps, in different grades. The investigator has before him a list of all the duties of the position, as definitely stated as possible. After each duty, is indicated the amount of time given by the encumbent of the position to that particular duty. When necessary this is followed up by a more detailed study of the work of the position at first hand. Forms used and concrete samples of the work are secured. The significance of all such forms used by the clerk is also indicated. With this as a basis the investigator may proceed to locate on the scale cards each of the duties of this position—find among the scale cards (exhibits) those which illustrate duties similar to those actually performed by the encumbent of the position in question.

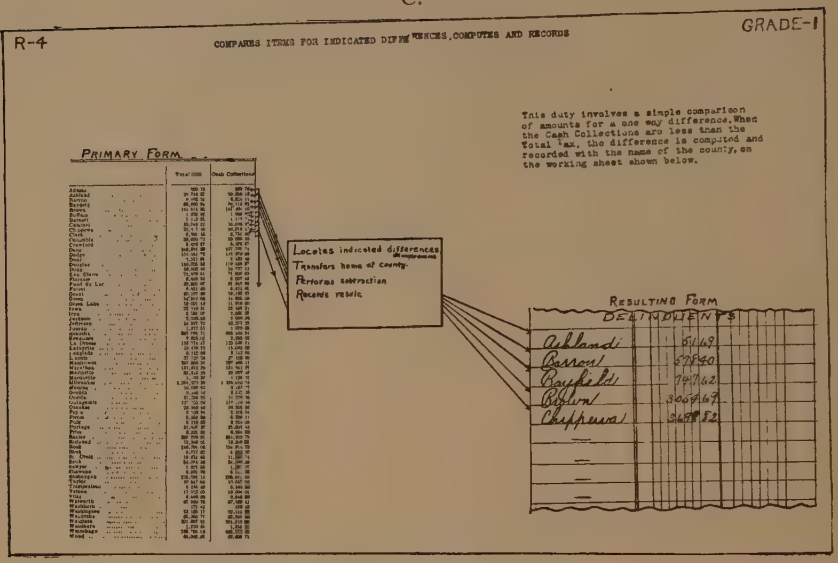
Let it be assumed that there are four types of duties connected with the particular clerical position in question. Each is given its proper grade, as indicated on the corresponding card in the scale. In connection with each is also given the relative amount of time devoted by the clerk to that particular duty. Thus each duty may be weighted as to the relative amount of time spent on such duty by the encumbent of the position. Then, in most cases, the position will be classified in that grade in which the major part of the duties of the position fall. The position is thus classified according to an objective standard. The results may be set down somewhat as follows:

Duties of Clerk A for Position X	Grade of Duty	Weight, or Proportionate Time Given Each Duty
Files cards alphabetically	1	.20
Tabulates income tax reports	2	.50
Checks rate computations	2	.20
Examines reports for inconsistencies	3	.10

CHART 2

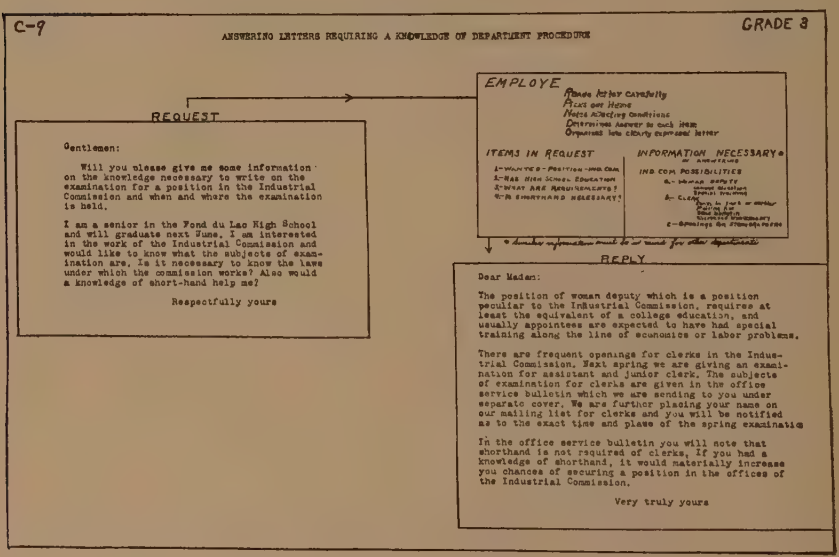
Illustrations from Clerical Performance Scale

C.


R-4
COMPARES ITEMS FOR INDICATED DIFFERENCES, COMPUTES AND RECORDS
GRADE-1
This duty involves a single comparison of amounts for a one way difference. When the Cash Collections are less than the Total Tax, the difference is computed and recorded with the name of the county, as the working sheet shown below.
PRIMARY FORM
Locates indicated differences
Transfers items of county
Performs subtraction
Records result
RESULTING FORM
DELINQUENTS
Ashland 5149
Barron 57840
Bayfield 79762
Burnett 3655 67
Chippewa 3459 82

Fourth Card from Recording and Bookkeeping Group

D.


C-9
ANSWERING LETTERS REQUIRING A KNOWLEDGE OF DEPARTMENT PROCEDURE
GRADE 8
EMPLOYE
Reads letter carefully
Files our Memo
Notes Affecting Conditions
Determines Answer to each item
Organizes into clearly expressed letter
ITEMS IN REQUEST
INFORMATION NECESSARY
REQUEST
Gentlemen:
Will you please give me some information on the knowledge necessary to write on the examination for a position in the Industrial Commission and when and where the examination is held.
I am a senior in the Fond du Lac High School and will graduate next June. I am interested in the work of the Industrial Commission and would like to know what the subjects of examination are. Is it necessary to know the laws under which the commission works? Also would a knowledge of short-hand help me?
Respectfully yours
REPLY
Dear Madam:
The position of woman deputy which is a position peculiar to the Industrial Commission, requires at least the equivalent of a college education, and usually appointees are expected to have had special training along the line of economics or labor problems.
There are frequent openings for clerks in the Industrial Commission. Next spring we are giving an examination for assistant and junior clerk. The subjects of examination for clerks are given in the office service bulletin which we are sending to you under separate cover. We are further placing your name on our mailing list for clerks and you will be notified as to the exact time and place of the spring examination.
In the office service bulletin you will note that shorthand is not required of clerks. If you had a knowledge of shorthand, it would materially increase you chances of securing a position in the offices of the Industrial Commission.
Very truly yours

Ninth Card from Correspondence Group

The duties of this position are now classified and weighted as a basis for salary determination, and also as a guide in the arrangement or weighting of tests or various parts of an examination for selecting clerks for this type of work.

STANDARDIZING THE GRADE OF DUTIES

In the determination of the grade of the clerical duties making up the Clerical Performance Scale described in the preceding sections, the writer had the assistance of nineteen employes from different departments of the state service, as follows: 5 chief clerks, 5 accountants, 2 statistical clerks, 2 filing and index clerks, 3 secretarial clerks and 2 professors of the university course in commerce. In securing the information desired the following steps were taken:

1. Each assistant was given five packs of small cards, each pack representing a type of clerical work (Correspondence, Filing, Statistical, Recording, and Examining and Certifying), and each card an individual typical duty under such type of work. The assistants were asked to arrange the cards of each pack in the order of difficulty, indicating such by a number on the card.
2. In step 2, the assistants were given an orderly arranged list of duties for each of the five types of work, resulting from the working up of the data secured from step 1. In this step, they were asked to proceed as follows: "Beginning with the lowest order duty, No. 1, carefully read down the series. When you come to a duty that you believe requires a distinctly higher type of ability, in its effectual performance, place a mark between it and the preceding duty. Continue in this manner to the end of each series." This resulted in several groups for each type of work, indicating grades of difficulty.
3. In step 3 the assistants were given a cross-section outline. They were also given slips on which were listed the duties for each group, resulting from working up the data of step 2. They were then asked "to locate each group of duties in the outline in such a way that in the vertical sections (were) groups of duties requiring practically the same degree of ability to perform, and in the horizontal sections duties of the same type, etc."

The clerical chart referred to on page 4 was made upon the basis of the study just described, and after modification resulting from conferences with those who assisted. The chart was further changed from time to time after the writer had made a more thorough and first hand study of the duties of different clerks.

THE FIELD OF CLERICAL DUTIES

With the series of illustrated clerical duties, as set forth in the clerical performance scale, divided as they are on the basis of kind

of work, and degree of difficulty, the field of interest of this study, as it exists in its broadest sense, is laid out. Along one line—the line of progress and promotion—are duties varying in degree of difficulty from the simplest office task to the one that is most complicated. Cross sections, along this line, show duties of practically the same degree of difficulty, but of different types. In the lower grades this difference between duties in the same cross-section may be illustrated by the difference between checking copied records for correctness, on the one hand, and filing cards alphabetically on the other hand. For the higher ranges of the scale, the contrast is illustrated by the comparison of the duties of the statistical clerk with those of the secretarial clerk.

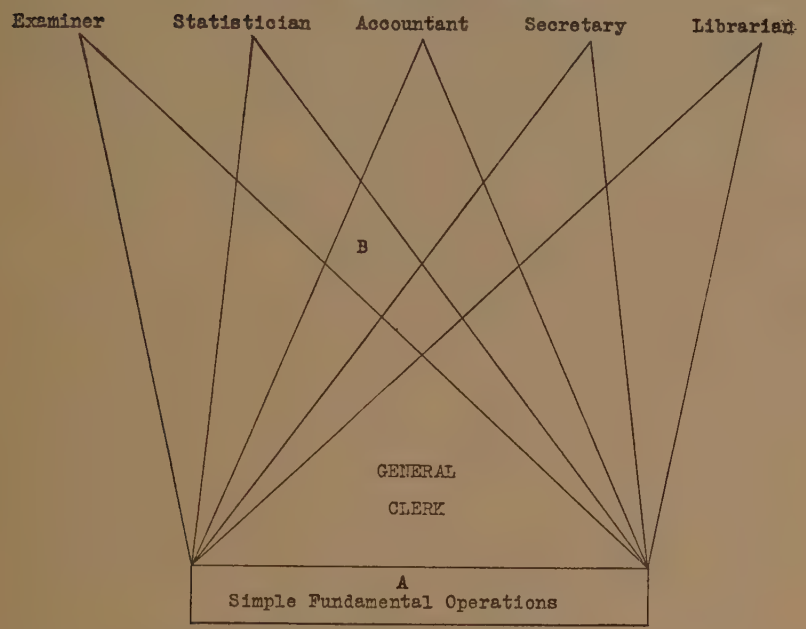
As one goes up the scale of clerical work, the differences become greater. Specialization becomes more apparent. The chart on page 13 is an attempt to suggest both the common foundation of all clerical work, and the divergence—as one goes up the scale. In the diagram are suggested five types of office work, each of a high degree of specialization, but resting on a common clerical basis. This common clerical basis includes such simple and fundamental operations as checking, comparing records, filing, easy tabulation, and mathematical computations. These operations in their simplest form are represented by that part of the diagram that is common to each of the five triangles—section A, the basis on which all the different lines of work rest. Even above this there is quite a section that is common to all five lines, but which becomes less as one goes up the scale. At the top of each triangle is found that specialization which is peculiar to the particular field suggested, as statistician, accountant, etc. Below this, however are sections of common ground. Much of what the expert accountant must know and possess in the line of ability must also be known and possessed by the expert statistician. The duties represented by section B might possibly be performed by a highly trained clerk who is neither a statistician nor accountant. Such duties might include the handling of mathematical data, and call for one trained in higher mathematics, and possessing special ability in that field. A high grade clerk in the field of library work, or filing and indexing, might be a failure in this mathematical work. As one goes down the scale there are less apparent differences between duties, and certainly less specialization. The question may well be raised, however, as to what degree differences in indi-

viduals affect their success in the different types of clerical work of the lower grades.

There is no distinct break in the line as we go from the lower to the higher grades of office work. At the lower end of the scale are duties of a routine nature, while at the upper end are those calling

CHART 3

COMMON "CLERICAL WORK" BASIS OF SPECIALIZED OFFICE POSITIONS



into play the highest functioning of the expert accountant or statistician. At such breaks as do occur the division into grades is naturally made. The work of the accountant overlaps that of the accounting clerk, and the work of the latter overlaps that of the lower grades of clerks.

FURTHER ANALYSIS AND SELECTION OF SAMPLINGS FOR THE TESTS

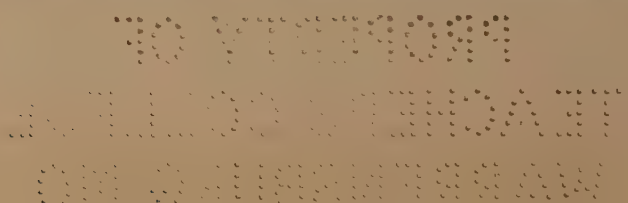
Thus far the consideration has been of the individual duty as the unit. Next an attempt must be made to break up the individ-

76878

ual duty into its parts or what may be referred to as its *fundamental operations*. Sorting or classification may be taken as an illustration of such a fundamental operation. An office boy is sorting cards into two piles according as they contain the names of American or foreign born. A round hole in the upper left-hand corner of the card signifies American born, and one in the upper right-hand corner, foreign born. A child of only a few years of age might, for a time, perform this fundamental operation of sorting quite satisfactorily. Even a non-English speaking child might be shown what to do, and go ahead with the sorting, having no idea as to what the whole thing was about. This particular illustration is taken to indicate the mechanical nature of this operation. In large offices where much of such work is to be done, holes are actually made as indicators of classification, and the cards are sorted by machine (Hollerith machine) instead of by a clerk, being done much more rapidly and accurately. This clerical operation is such as requires only a low degree of intelligence—just about enough to get and keep the necessary mental set for the operation. The maintenance of this mental set, and an accuracy and alertness of perception, is what determines to a large degree the efficiency of the boy's work, aside from a certain manual dexterity in handling the cards, and a minimum of motion in performing the manual part of the operation. Here is a clerical operation at the very bottom of the scale. Even though instead of holes in the card, a check had been made after the appropriate designation—"American Born" or "Foreign Born"—the operation is not raised to any material extent.

At the other extreme, in the performance of sorting, is the work of the expert accountant. He goes into the offices of a business concern to install a new accounting system. One of his big problems is to establish a classification of accounts that fits the business in question. After studying the situation, he formulates a tentative scheme of classification, and proceeds to arrange (sort) the various items (expenditures, etc.) in an experimental fashion. He is sorting or classifying when he is doing this, but it is classification in one of its highest forms. With the office boy and his job not much mentality is required, no experience, and no knowledge of what it is all about. With the accountant, years of training are required, a high degree of mentality, and a wide range of knowledge.

The requirements in each case may be summarized as follows:



1. In the case of the office boy:

   Facility of performance: Making the separation into two piles. An accuracy of discrimination, and a quickness and accuracy of reaction, in which the manual factor is quite important.

   Intelligence: Only a very low grade required.

   Knowledge: Not even a comprehension of the meaning of the process is necessary.

2. In the case of the expert accountant:

   Facility of performance: Speed, in terms of which the work of the office boy might be measured, is almost a negligible factor.

   Intelligence: A very high degree of intelligence is required. Analytical ability, initiative, and an expertness at systematizing, as well as ability to think in abstract terms, are necessary.

   Knowledge: Extensive knowledge of the principles of accounting, and a knowledge of the business in connection with which he is working, are absolutely essential.

A study of typical clerical duties shows the possibility of arranging, between these two extremes, a scale including sorting or classification operations of some twenty-five different degrees of difficulty. Here, as well as in connection with other fundamental clerical operations, the important factors for the lower grade positions are rapidity and accuracy in rather simple processes; while for the higher positions these factors decrease in importance, and the exercise of the higher mental functions, together with a widening range of knowledge, come to the front as essential.

A careful study of clerical work, especially as illustrated by the clerical performance scale cards, resulted in the selection of six fundamental operations,[1] with varying possible ranges of difficulty within themselves. Those determined upon as practically covering the entire cross-section of clerical work are as follows:

1. Checking of material for correspondence with copy or similar material, with which it is compared for sameness.
2. Selection of data on the basis of comparison.
3. Locating in an alphabetical list.
4. Classification or sorting.
5. Tabulation or selection and transfer of data.
6. Mathematical computations.

These are not set forth as all-inclusive but as giving the basic operations involved in practically all clerical work. In actual clerical duties we do not, as a rule, find these fundamental opera-

[1] See Appendix, page 85, (4).

tions in their simplicity, but rather in various combinations as will be shown directly.

The clerical duty which is here taken as an illustration is that suggested in the chart on page 10. The generalized statement of the duty is: "Compares items for indicated differences, computes and records." In the particular form (exhibit) used as illustrative, the clerk examines the reports of total taxes and cash collections as they come to the office, and lists on a separate sheet the names of those counties that are delinquent, together with the amount of such delinquency. Three different operations are involved in the performance of this duty—three of the fundamental operations set forth above. They are as follows:

1. Selecting data on the basis of comparison.
   The clerk goes down the list of counties, comparing the figures in the two columns, and selects for further consideration those counties for which the two figures are not the same.
2. Performing simple arithmetical operations.
   This process of subtraction gives the delinquency.
3. Transferring data from one form to another or tabulating.
   In the transfer, the name of the county and also the difference are transferred from the sheet where the subtraction was made to the "resulting form"—shown at the right.

This shows for a rather simple duty how three fundamental operations enter into its make-up. In a similar fashion any clerical duty may be analyzed into its fundamental operations. In many duties, of course, a much more elaborate combination of these operations is found. In most of the lower grade clerical duties the order in which the fundamental operations follow one another is more or less fixed, and easily learned. As one goes up the scale, however, the determination of their order is really the thing that makes the greatest demands upon the clerk. In many cases the results of each operation is the determinant of what the next operation shall be. Here, in the higher clerical duties, independent judgment enters in at many places in determining which operation is to follow. This, then, is what largely distinguishes the higher types of clerical work from those of a simple routine nature.

In this study the point is now reached where the field may be limited for investigation. The attempt was here made to devise tests for determining facility in performing the simpler clerical duties, where extensive knowledge or high intelligence is not necessary.

CHAPTER II

THE SIX TESTS AND THEIR ADMINISTRATION

The sampling tests used in the battery were based on the six fundamental operations enumerated in the previous chapter, which briefly restated are as follows:

1. Checking for sameness.
2. Comparing.
3. Locating alphabetically.
4. Classifying and sorting.
5. Tabulating.
6. Computing arithmetically.

DESCRIPTION OF THE TESTS

The six tests making up the battery, samples of which appear on pages 19 to 23 inclusive, were printed and gotten out in folder form, all tests except the sixth taking two pages. In the construction of the tests and the folder, an effort was made to keep the content as nearly as possible like that which might be found in the material of an office. The suggestions for content were rather naturally taken from the operations found in the offices of the state service. The whole attempt was to create an atmosphere approaching that met in an actual office. To this end the statement appearing on the cover (page 18) was made. The word "test" or "examination" does not appear in the folder, "exercise" being used in place of either of these designations.

Exercise 1

In test one (exercise 1) of the battery (page 19) the checking or comparison is, as will be seen, for sameness of copy and original. The items—name, address and number—are such as, in part or in whole, are found in most of this sort of checking in ordinary office procedure.

Exercise 2

Here the comparison is for relationship rather than for sameness, and the content is limited to numbers. Operations of this type

CLERICAL EXERCISES

Number .....

Position applied for..... Date .....

DIRECTIONS

The clerical exercises in this folder take about an hour to complete. For the sake of uniformity you are asked to begin and stop work on each exercise when told. It is not expected that you will finish any of the exercises in the time allowed. Ask no questions. You will find all necessary directions at the beginning of each exercise.

Do your work carefully. Imagine that you are doing actual clerical work in an office under regular office conditions. Keep uppermost in mind the idea of accuracy.

Waste no time. Work as rapidly as you would if you knew that other work was being held up until what you are doing is completed.

In each of the five exercises the directions have already been carried out for the first two or three cases as samples of what you are to do with those following. The samples appear in *italic*.

When directed copy the following sentence, in your best handwriting, using the space below.

"The clerk who keeps an orderly desk uses much the same sort of ingenuity and method used by the manager who keeps an orderly business."

Do not open folder until told to do so.

are frequently found in connection with the work of a bookkeeper or accountant, and in the examination of various reports, especially in public service where such comparisons are often required. In this exercise (test) exact subtraction or division is not called for. In most cases, a glance at the numbers ought to be sufficient to determine the reaction to be made.

SAMPLE FROM PAGE 2 OF FOLDER

EXERCISE 1

OUT-OF-STATE TAX LIST

Directions:-Check at the right as indicated, each line where the name, address or amount is not exactly the same in the copy as in the original.

ORIGINAL			COPY			Check Here
Henry Reeves	*Carey, Ohio*	*$18 21*	*Henry Reeves*	*Carey, Ohio*	*$18 21—*	
Agnes L. Dwight	*Andover, N. J.*	*6 42*	*Agnes L Wright*	*Andover, N. J.*	*6 42—✓*	
Cecile Fennelp	*Belfast, Tenn.*	*10 09*	*Cecile Fennelp*	*Belfast, Tenn.*	*10 09—*	
John Dowling	Blake, S. C.	25 83	John Dowling	Blake, S. C.	25 83—	
Sarah Hanson	Ashland, Ohio	1 17	Sarah Hanson	Ashland, Ohio	1 07—	
Mrs. A. L. Smith	Ely, Nevada	4 77	Mrs. A. L. Smith	Ely, Nevada	4 77—	
D. H. Clark	Ellegan, Mich.	18 94	D. H. Clark	Allegan, Mich.	18 94—	
Sam Potter	Monroe, Wis.	26 14	Sam Potter	Monroe, Wis.	16 14—	
Kate Maloney	Euclid, Ohio	91 49	Kate Maloney	Euclid, Ohio	91 49—	
Mrs. I. O. Kay	Austin, Pa.	11 07	Mrs. I. O. Kay	Austin, Pa.	11 07—	
Rev. A. R. Green	Glasgow, Mont.	2 95	Rev. A. R. Green	Glasgow, Mont.	2 95—	
Edna Moore	Ada, Minn.	54 93	Ed. N. Moore	Ada, Minn.	54 93—	
Anna Randall	Sidney, N. Y.	13 03	Anna Randall	Sidney, N. J.	13 03—	

Material continued through page 3

SAMPLE FROM PAGE 4 OF FOLDER

EXERCISE 2

POPULATION OF CITIES AND PATRONS OF A PUBLIC UTILITY

Directions:—Where the number of patrons is more than half the population of the city place a check after the dash as indicated.

CITY	POPULATION	PATRONS	CITY	POPULATION	PATRONS
Alton	*7 322*	*5 435—✓*	Etna	14 378	7 947—
Abbot	*26 253*	*11 648—*	Eton	3 006	1 111—
Alberta	*4 201*	*2 980—✓*	Everson	1 336	610—
Amherst	3 648	1 210—	Excelsior	31 092	14 015—
Ardley	11 743	6 098—	Exeter	7 701	4 405—
Arlington	11 485	5 014—	Fairveiw	2 091	876—
Armada	4 048	2 001—	Fayette	10 155	4 465—
Artesia	7 483	4 714—	Forbush	6 335	3 100—
Atlanta	22 514	12 601—	Fox Lake	4 809	2 304—
Auburn	5 109	2 121—	Freedom	12 362	7 102—
Bagely	27 104	9 195—	Fulton	31 017	15 117—
Bath	19 352	15 921—	Gacho	2 995	1 786—
Bayview	2 190	1 768—	Green	3 102	1 456—
			Grinnell	11 830	5 104—

List continued through page 5

Exercise 3

This is a straight alphabetizing test (exercise) calling, of course, for a knowledge of the alphabet. The operation is quite similar to that of a clerk locating cards in a file, and was devised with that in mind. When the examiner finds the space between two names (cards) where the name to be located alphabetically belongs, a pencil mark or line is drawn in the space, and at the end of the line

SAMPLE FROM PAGE 6 OF FOLDER

EXERCISE 3

NAMES TO BE FILED ALPHABETICALLY

Directions:—The names listed on the next page are in alphabetical order. Insert in that list, in its proper alphabetical place, each of the 50 names given below. For convenience, insert the number of the name rather than the name itself. The first two names, ***Becker*** and ***Baer*** are so inserted, by number, as an illustration.

1	*Becker*	26	Blew
2	*Baer*	27	Bergold
3	Bogart	28	Barter
4	Bibbs	29	Beebe
5	Bauer	30	Barry
6	Bachmann	31	Berze
7	Blau	32	Banner
8	Bayley	33	Bishop

50 names on this page

SAMPLE FROM PAGE 7 OF FOLDER

NAMES FROM A SECTION OF FILES

Baas	Barkpess	Beaulage	Berigan	Black
Babcock	Barland	Beaumont	Berillus	Blackburn
Babilijon	Barlow	Beauverd	Berkan	Blackler
Bach	Barnard	Beaver	Bernardi	Blackmer
Bachelder	Barnes	Bechtel	Bernstein	Blair
Bacher	Barney	Bechter ——1	Berschens	Blake
Bachmeyer	Barnwell	Beckett	Berthold	Blakeman
Bacilicom	Barrette	Beckler	Bertholdt	Blakemore
Backus	Barrigen	Beckman	Besch	Blanchar
Bacon	Barrimage	Beckner	Best	Blanchard
Badami	Barron	Beckter	Besta	Blanche
Badeaux	Barrows	Beckwith	Bestan	Blanco
Badger	Barruwell	Beecroft	Bestur	Blaney
Badke ——2	Barsantee	Beeken	Bethge	Blaschka
Baessee	Barsness	Beers	Betlach	Blawusch
Baff	Bartec	Beglinger	Betz	Blazek
Bagley	Bartel	Bebnke	Beuthin	Bleck

List fills entire page

is indicated the number by which the name is designated in the list on the preceding page of the folder.

The physical reaction, to be sure, is not the same as in actual filing, neither is the mental operation just the same; but the essential factor—that of perceiving the letter make-up of a word—is quite the same as it appears in the everyday work of a filing clerk. (See discussion of alphabetizing on page 24.)

Exercise 4

This test (exercise) of classification represents a large number of clerical duties. In the selection of the material of this test the endeavor was made to find such as would be about equally familiar to all who might be applicants for clerical positions. In this test a quickness and surety of association are called for. As will be seen, the operation called for in the performance is not the easiest that might have been introduced, but is certainly such as any applicant should be able to perform with a reasonable degree of rapidity and accuracy.

SAMPLE FROM PAGE 8 OF FOLDER

EXERCISE 4

INSTITUTIONAL ACCOUNT

Directions:—Indicate with proper number (*as illustrated*) those items on the list below that may be classified under one of the following heads. Only part of the items can be thus classified.

Classification	*Use these numbers*
Food	*1*
Furniture	*2*
Clothing and clothing material	*3*
Personal services and wages	*4*

50 lbs of bacon.....*1*	5 dish pans.....
10 bottles ammonia.....	100 lbs. sugar.....
kitchen table.....*2*	bookcase.....
plumber, 4 hrs.....	25 cedar posts.....
1 large arm chair.....	½ doz. pie tins.....
day labor, 9 days	hay rack.....
100 lbs. pork	harness.....
small table.....	6 milk pails.....
frying pan.....	4 reels barbed wire.....
gardener's wages.....	hall rack .....
new barn door.....	12 pr. overalls.....
postage stamps.....	½ doz. hoes.....
3 sacks flour.....	5 doz. boys' caps.....

List continued through page 9

Exercise 5

In many offices, and especially in connection with statistical work, there is to be found tabulation similar to that incorporated

in Test 5. In this test there is a rough sorting operation, but the test as a whole is supposed to emphasize accuracy in picking out and transferring numbers. From one type of work to another calling for this operation, the content is apt to be similar with probably a considerable variation in form.

SAMPLE FROM PAGE 10 OF FOLDER

EXERCISE 5

CROP ACREAGE BY COUNTIES

Directions:—Fill in the blanks in the table on the following page, using data found below. Make neat figures and keep them in allignment.

ALBANY		DESOTA		JACKSON		NEWTON	
Corn	*23 157*	Rye	8 879	Rye	66 765	Barley	34 610
Barley	*1 546*	Wheat	798	Corn	44 454	Rye	4 220
Oats	*18 231*	Oats	4 985	Oats	3 265	Corn	4 210
Rye	*625*	Barley	5 465	Barley	528	Oats	4 091
Wheat	*26 937*	Corn	1 298	Wheat	971	Wheat	750
BARBER		GARRET		LAWRENCE		NORTON	
Oats	5 239	Wheat	9 108	Barley	6 263	Rye	29 468
Wheat	4 367	Oats	7 565	Rye	8 564	Corn	5 120
Corn	1 586	Barley	675	Wheat	24 378	Barley	540
Barley	2 978	Corn	5 768	Corn	1 337	Wheat	1 254
Rye	22 754	Rye	8 915	Oats	88 421	Oats	5 908
BARRON		GASTON		LEWISTON		ORANGE	
Barley	65 322	Oats	7 426	Wheat	921	Corn	2 220
Oats	978	Corn	6 237	Corn	73 225	Barley	6 141
Wheat	546	Barley	56 190	Oats	1 234	Oats	404
Rye	4 234	Rye	897	Rye	7 662	Rye	189
Corn	456	Wheat	675	Barley	4 386	Wheat	9 682

Data fill entire page

SAMPLE FROM PAGE 11 OF FOLDER

ACREAGE DEVOTED TO BARLEY, RYE, AND CORN

County	Barley	Rye	Corn	County	Barley	Rye	Corn
Albany.....	*1546*	*625*	*23157*	Jackson.....			
Barber.....				Lawrence...			
Barron.....				Lewiston...			
Chase.....				Lowell.....			
Clay.....				Lyon.....			

Table fills entire page

Exercise 6

The last of the six tests is one of alertness and accuracy in performing the simple arithmetical operations of addition, subtraction, multiplication and division. Their arrangement in mixed order

calls for ability to shift quickly from one operation to another, as is often required in actual clerical work, and which in itself sometimes leads to slower speed and inaccuracy.

SAMPLE FROM PAGE 12 OF FOLDER

EXERCISE 6

Directions:—Perform the following arithmetical operations. **Proceed from left to right** completing the eight problems in the top row before beginning on the second, the second eight before beginning on the third row, etc.

Add	Subtract	Multiply	Divide	Add	Subtract	Multiply	Divide
27 32 96 155	348 293	164 9	8)6728	19 27 95	407 198	297 7	6)4068
42 58 69	901 232	478 6	4)2068	73 64 38	726 459	586 8	9)8496

IN LOWER CORNER OF PAGE 12—FOR RECORDING RATINGS

Exercise	Amount	Errors	Score	Weighted Score
1. Checking Copy				
2. Comparison				
3. Filing				
4. Classification				
5. Transfer				
6. Computation				
Penmanship				
FINAL SCORE				

REACTION RATHER THAN KNOWLEDGE TESTS

These tests are seen to be tests of quickness and accuracy of reaction. They are not designed as tests of knowledge or intelligence. A certain minimum amount of knowledge and a minimum intelligence are of course required to read and comprehend the directions, but these minimum limits are such as any grammar school graduate should have attained. An attempt was made to keep the directions as simple as possible, and to choose subject matter or content that would be simple enough to give the ordinary applicant for a clerical position neither advantage nor handicap in

comparison with the others. Of course this ideal could not be wholly realized.

In the tests the operations are largely simple or single operations. In other words, they do not, in general, call for a combination of the "fundamental operations." The discovery of a candidate's alertness and accuracy in these simple fundamental operations was what was attempted. The thought was that a combination of this test with a general intelligence test might be worked out for the selection of clerks for positions where not only speed and accuracy in clerical performance are desirable, but where intelligence plays a considerable part in determining the order of the clerical operations—the plan of procedure.[1]

The test in alphabetizing furnishes a concrete illustration of what was attempted in formulating the tests. In many civil service examinations for clerks, there has been a filing or alphabetizing test. A typical sample is here given.

Arrange in correct alphabetical order the following names:

Henry K. McManus
Peters, Brown & Company
John M. Grant
MacBride, Howard & Company
B. J. Peters
Warren & Connover, etc.

In such examinations the candidates had practically all the time they needed, and no one knew whether a certain candidate took five or twenty minutes to arrange the names. The candidate must know whether to put names beginning with Mc, as in McManus, before or after those beginning with Mac, as in MacBride, as well as in relation to the other Ms. He must also know whether B. J. Peters comes before or after Peters, Brown & Co. This is illustrative of the type of thing determined in the filing questions in such an examination. They are quite distinctly knowledge questions.

In the filing test of this study, the point of view was taken that for the routine position of filing clerk, or in the case of the clerk devoting a large part of the time to such filing, accuracy and speed are the essential factors, and that the very few problems of order, such as those suggested above, can be quickly learned by the average clerk in a very short time. The essential thing, it

[1] See page 16, paragraph 3.

seemed, is the ability to quickly perceive the letter make-up of a word. This done accurately and quickly constitutes the foundation of the work of a good filing clerk. With this in mind, the test was devised in the form in which it appears. For this reason, only one letter of the alphabet was taken and initials were omitted. In many large offices cards to be filed are first arranged or grouped by first letters, and then arranged or filed according to the interior letter make-up of the names. This will serve to illustrate the point of view taken in devising the tests.

METHOD OF ADMINISTERING THE TESTS

In giving the tests in regular civil service examinations, an effort was made to do nothing to make the candidates nervous. The watch was kept out of sight in so far as possible, and the directions to stop one test and begin on the next were given in an ordinary tone of voice, and in the form of a statement rather than by saying "Start!", "Stop!". As will be seen from the copy of the front page of the folder, shown on page 18, an exercise in penmanship was given first as a sort of shock absorber.

A definite time was allowed for each test as follows:

Test 1	5 minutes	Test 4	5 minutes
Test 2	5 minutes	Test 5	7 minutes
Test 3	8 minutes	Test 6	3 minutes

In the large clerical examinations held in the state capitol at Madison, the tests were given by the writer, but to the smaller groups in the various centers about the state, local examiners gave the tests, following the directions given on page 83 in the Appendix.

THE PRELIMINARY TRYOUT

The tests, in mimeograph form, were first tried out in connection with a group of 99 candidates in a regular civil service examination for beginning clerks. They were given before the regular examination, and the announcement was made that the results would not in any way affect the candidates' grades on the civil service examination. On the basis of this tryout the time allowances were determined for the different tests, and several minor changes were made.

CHAPTER III

THE CRITERIA

With the sampling tests actually constructed, the preliminary trial out of the way, and suggested corrections incorporated in the tests, the next problem was that of determining the selective value of the tests for clerical workers. Inasmuch as the tests were based on the results of a job analysis, it was thought that the most significant results could be obtained by giving the tests to several groups of clerks of known efficiency, each of the several groups representing a different type of work, and all employes in any one group doing the same type of work, under the supervision of one person from whom ratings comparable for all in the group might be obtained.

Groups of sufficient size meeting the above conditions were not to be found in the state service. It was therefore necessary to look outside for such criteria. The writer was fortunate in this connection in securing the coöperation of several large and nationally known business concerns. As a result of this coöperation the test was given to groups of clerks in the offices of the following business concerns: a large department store; a clothing manufacturing establishment; an insurance company; two public utilities; and a large manufacturer of electrical equipment.

The test was also given to several other groups, not coming under the specifications set forth in the preceding paragraphs, but on which helpful data were secured. Among these groups were the following:

Candidates in civil service examinations in the state of New York—about 900 in number.

Candidates taking the civil service examination for clerks in the state of Wisconsin.

Stenographers and clerks in the Extension Division of the University of Wisconsin.

Copy clerks in the office of the County Recorder of Milwaukee County.

Students in the Madison High School—about 200 in number, mostly commercial students.

The following list shows the different types of employes, in

the various business concerns mentioned, to whom the test was given, and the number in each group. The code letters that will be used throughout this study in designating these groups are also given.

Code Letter of Group	Title of Clerks[1]	Number in Group
A	Billing Clerks	27
B	Billing Clerks—Elliott Fisher Operators	49
C	Dividend Clerks	21
D	Auditing Clerks	24
E	Order Clerks	15
F	Pricing Clerks	22
G	Pay-roll Clerks	15
H	Pay-roll Clerks	15

DIFFICULTIES IN SECURING ADEQUATE CRITERIA

The chief difficulty in securing criteria was that of finding groups of sufficient size meeting the conditions set forth. It is recognized that the resulting correlations would have been much more reliable had the groups been larger. On the other hand, it should be noted that the test data from a number of such small groups are in some respects superior in value to those which might have been obtained from an equal number in one group. The variability of the test results from group to group is brought out, and greater significance is attached to those results that seem to hold for all groups than had similar results been observed in one large group. Of course the ideal condition would be to have a number of large groups. In so far as an individual test may be used to advantage in selecting general clerical workers, the results of the test, as given to diversified groups, are especially significant. Diversification further serves the purpose of indicating the special value of individual tests as selective agents for different types of work.

The writer was able to give the test personally to only a part of the groups, but he has no reason to believe other than that it was administered with care and interest to each of the groups, according to the directions given in the Appendix, page 83. Probably not all groups worked under equally favorable desk arrangements, but such differences can hardly be avoided. It is further possible, of course, that the actual time allowed for any test might vary a few seconds from that indicated. Although this might raise or lower

[1] See Appendix for descriptions of duties of clerks.

the average performance of that group, it would hardly be expected to have any very noticeable effect upon the correlations of that group.

A WORK ANALYSIS FOR EACH GROUP

In the case of each group to which the test was given, a written analysis of the work of the position was secured, with samples of the forms used in the work. This was supplemented by such further information about the work as seemed necessary. On the basis of this information an endeavor was made to determine the significance of the tests in connection with the different types of work represented by the groups. This was done by comparing the fundamental operations involved in the different types of work, with the fundamental operations serving as the bases of the tests. In the Appendix will be found a detailed description of the duties of each group of clerks to whom the test was given.

CRITERIA RATINGS

The supervisor of each group was asked to rank the clerks in that group on the basis of quality and quantity of work done. The nature of the instructions follow:

"Rank in order of efficiency in work done, the clerks taking the test. Disregard personality, taking into account quality and quantity of work only.

"Give the one ranking highest 1, the second 2, and so on. Do not give two the same rank. If it is difficult to distinguish between several clerks, you may put a bracket about their names to indicate this, but give each a different rank—that which you consider the most probably correct."

A study of the resulting correlations showed that in these ratings the factor of length of employment was playing a large part, as of course might be expected. It was decided to ask for another rating. The gist of this second set of directions for rating is given in the following paragraph.

"If all the employees started in at the same time, without any previous experience in your office, how would you expect them to rate up at the end of a year, as a result of their relative native aptitude for the particular work of your office?"

In the case of the largest group (49 clerks) the supervisor this time gave a percentage rating to the clerks. She was asked to arrange the clerks in a rank order, then to give the highest 100 per cent, and to the others marks that would indicate their relative differences.

For six of the groups then, there were criteria based both on efficiency of work, and on aptitude.[1] In Table I are shown the correlations between each of these two ratings and length of service, as well as between the ratings themselves. Only four groups are included in the table, as they were the only ones for which the length of employment was secured.

The rank scores obtained were transmuted into sigma deviations on the assumption of normal distribution.

TABLE I

Correlations[2] between Ratings on and		Aptitude and Efficiency	Length of Service and Efficiency	Length of Service and Aptitude
Groups	A	.49	.16	.08
	B	.55	.79	.35
	C	.77	.85	.47
	D	.80	.72	.40

At this time attention will only be called to the fact that the efficiency ratings for Groups B, C, and D showed high correlations with length of service, the correlations being very materially less between aptitude and length of service. Even in the aptitude rating, however, the influence of length of service appears to be quite marked in the case of three groups. Group A, showing the lowest correlation in both cases with length of service, is a group where the range and difference in lengths of service are comparatively small.

THE VALIDITY AND RELIABILITY OF CRITERIA

All who have worked with tests and standardization realize the uncertainty of criteria, and know of the problems arising in connection with their use. Experience in securing and interpreting efficiency ratings on employes of all types, as secured from many supervisors of such employes, has impressed upon the writer the

[1]This number is increased to 7 if what the writer considers as approximately an aptitude rating is included. This is the ratings for group A, which although not made out according to the directions given above were, it seemed, quite surely based on observed aptitude, as indicated by a conference with the supervisors. The other set of ratings for this group was based entirely on the bonus received.

[2]No attempt is made in the presentation of this study to give the probable errors. This is because of the large number of correlations involved and the varying sizes of the groups considered in the study. Those interested in working out the probable error for any correlation will find all the necessary data in the tables presented. The number of cases in each group will be found on page 27.

fact that many factors operate to reduce the reliability of judgments of human ability and worth. Men like Thorndike and Rugg have treated this subject most helpfully, and each year it is found on the program of the National Civil Service Assembly. It seems inexhaustible, and yet its repeated discussion emphasizes its importance. Rating scales have been made, and are being experimented with; and yet skepticism prevails as to the efficacy of any so-called rating system. In the face of all the excellent discussions of this subject with the lists of factors that tend to affect the reliability of human judgment in this field, the writer is venturing to offer another such set of pertinent factors. This is done with no idea of its being an improvement over those classifications that have been already offered by others, but because it makes an easy and satisfactory approach to a point which it is desired to emphasize in this connection.

FACTORS TENDING TO AFFECT THE RELIABILITY OF RATINGS

The factors presented in the following paragraphs, although of general applicability, will be considered especially as they apply to clerical workers.

Personal Prejudice is no doubt the first factor that comes to the mind in the consideration of this subject. Its effect in distorting judgment of true human worth is too well known to need discussion here. Sometimes it is consciously present when a rating is given, and sometimes unconsciously.

Superficial Judgments concerning individual worth are usually based upon superficial information or a lack of the facts which should be possessed in order to pass reliable judgment. In the operation of this factor, true worth is often penalized to the advantage of the employe who is an artist at bluffing. In this situation the employe whose work, by chance or intention, is brought especially to the notice of the employer is apt to receive credit or penalty that is out of proportion to that deserved.

General Intelligence or Adaptability is often quite inseparable, in the mind of an employer, from worth along any particular line, even when a low degree of intelligence is all that is required. Thus, in the case of a clerk of high intelligence, doing simple routine work in only a fairly satisfactory manner, the tendency is, through the operation of this factor, to rate her above what she deserves on the basis of her actual work on the job. Many employers prefer clerks

of a high degree of intelligence even for routine work. General intelligence is closely associated with adaptability or ability to learn. In fact those factors are important indices of general intelligence. Yet, on the other hand, an increasing degree of intelligence may mean less satisfaction with routine work, and perhaps a lower grade of work as time develops the routine into a painful boredom.

Variability in the Judgment of the Supervisor is another factor that operates to affect the validity of ratings. Edgeworth, the English statistician, brings out this point clearly and concretely in his discussion of the causes of errors in examination ratings.[1] He attacks the problem with considerable refinement, and with the use of statistical instruments. The presence of this factor in judgment ratings has been especially demonstrated in connection with the marking of examination papers. The writer discussed this subject several years ago in connection with a thesis on "Grades and Grading,"[2] in which the whole subject of reliability of examination marks was analyzed.

Emphasis upon Errors or Mistakes in Work is a factor present in the determination of ratings given employes. Closely associated with the preceding factor of variability, mistakes in work often have a varying effect upon the supervisor's judgment, and thus his rating of the employe. It is especially marked, immediately after an annoying mistake made by the employe, and wears off with the passing of time. What we may call the average degree of annoyance with errors in work, seems to vary quite apparently from one employer or supervisor to another. Errors which one supervisor might scarcely notice, are with another exceedingly annoying, and very materially affect the rating of the erring employe. Emphasis upon mistakes of various types is the resultant of two causes: the idiosyncracies of annoyance of the particular supervisor, and the relationship of errors to the general organization and efficiency of the office.

Lack of Standard is a factor commonly operating to make impossible what we might call a truly reliable rating. This leads to the question as to what a reliable rating really is and the necessary conditions for such a rating. Suppose it were possible to determine accurately to just what degree each clerk in an office possessed

[1] Edgeworth, F. Y. "The Element of Chance in Competitive Examinations." *Journal of the Royal Statistical Society*, 1890, p. 645.

[2] Ruggles, Allen M. *Grades and Grading*. Master's Thesis, Teachers College, Columbia University, 1911.

any quality desirable in a good clerk. Further, suppose that six abilities—characteristics—were determined upon as entering into the make-up of a good clerk. With such reliable measures of these six qualities for each and every clerk, would the problem be solved? There still remains the problem of the relative emphasis to be placed upon each of these six qualities. As some clerks would possess more of one quality and less of others, the resulting rating would vary materially with any change in the emphasis or weighting of these known qualities. In other words, could one have just what one wanted in a clerk, just what *would* or just what *should* one want? What is the ideal combination of qualities?

THE OBJECTIVE OR PURE OUTPUT CRITERION

This leads to the question of what one really should desire in a reliable criterion to be used in checking up the value of a test of the nature of the one of this study. Is a record of pure output the best that can be obtained? Such a rating, it may be said, would surely be objective, and therefore ought to serve the purpose well. It may be suggested, however, that personality is an important factor in determining a clerk's value to an office—that an employe's attitude is certainly to be considered in connection with any office where morale is looked upon as an element of good management. But we may dispose of that suggestion for the present by saying that such factors, though essential, may be taken care of separately, and that such a test as the clerical test of this study is not devised for measuring such qualities or traits as attitudes. With this suggestion disposed of, a purely objective rating, such as is found in a bonus record for an office, will be considered.

BONUS RECORDS AS CRITERIA

A bonus is based upon just the objective thing for which we are looking. It is based upon pure objective work records. A bonus is figured according to some plan adopted by the concern in question. The following is illustrative of one system, which will be considered in the discussion immediately following: 8 cents each for the first hundred in output; 12 cents each for all over a hundred; 1 cent deduction for each error.

Here again the same problem confronts us—emphasis upon errors. A one cent deduction for errors indicates quite a different

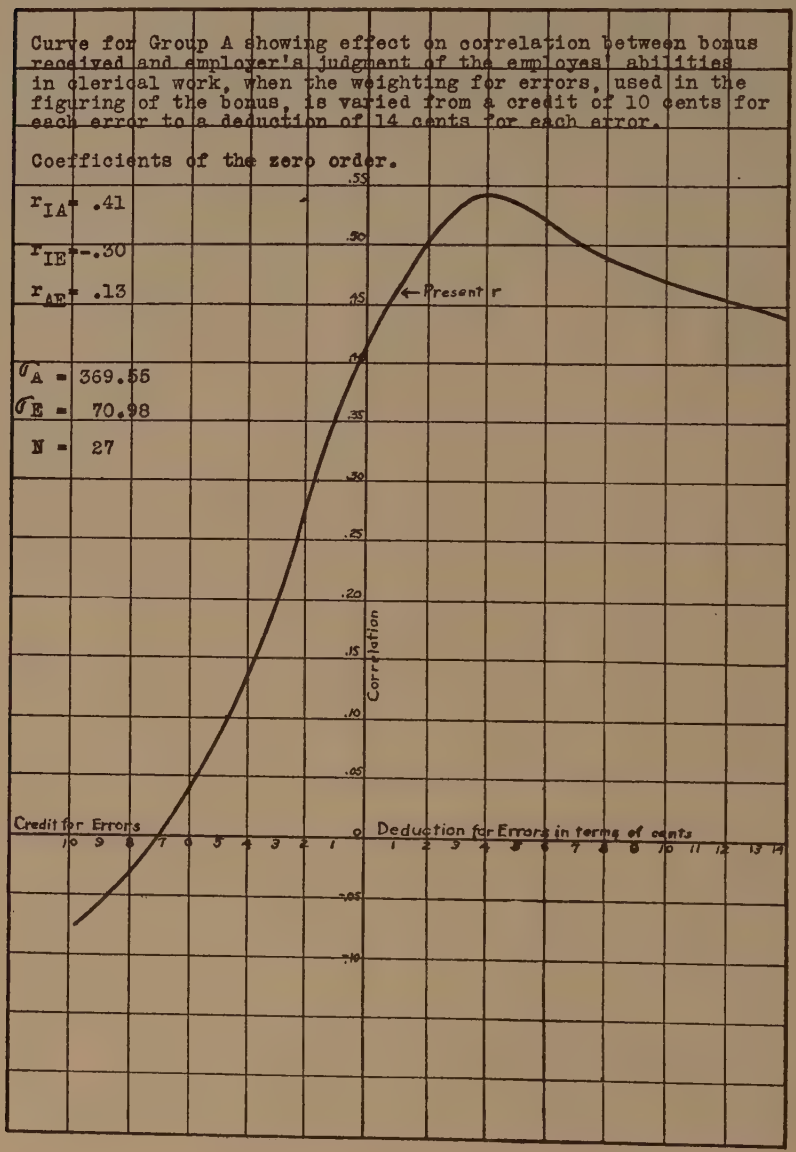
emphasis upon errors than would a ten cent deduction, and in turn might be reasonably expected to have an effect upon the number of errors made. A ten cent deduction would seem to offer a greater incentive for accuracy than a one cent deduction. How has it been determined that a one cent deduction for errors is best, if it is the best? No attempt will be made to answer this with any degree of assurance, but a safe guess would be that, in a majority of cases, it is determined in about the same way as deductions for errors are determined upon in the grading of school papers or in civil service examinations. In one concern using the bonus system an officer made the statement: "Possibly we do not penalize enough for errors." And one would hardly expect to find a concern where any real sense of surety in the method used would be expressed. This pushes the investigator back to the problem of finding the best weighting or deduction for errors, in figuring the bonus. The objective criterion has vanished. It may be suggested that there is no need of taking the dollars and cents rating, where one cent has been deducted for errors; that it would be better to go direct to the work records themselves. But such procedure only shifts the burden of assigning weights for errors from the employers to the investigator—perhaps a case of "from the frying pan into the fire."

VALIDITY OF A GIVEN WEIGHT FOR ERRORS IN DETERMINING BONUS

In the case of one concern the writer received two sets of ratings. One was the bonus received, and the other the estimate of the supervisors—a rating that apparently would be very satisfactory to the supervisors and office manager, as representing the clerks' fitness for that particular kind of office work in which they were engaged. In the minds of these supervisors the ratings were based upon the clerks' possession and display of ability as billing clerks. For this group the curve represented in Chart 4 shows the variation in correlation (between the bonus ratings of the clerks and the ratings given by the supervisors) resulting from the use of different deductions for errors. One cent is the amount actually deducted by the company in figuring the bonus. It will be seen, however, that if the office manager believes that the ratings of the supervisors are more nearly representative of the value of the clerks to the concern, he could better penalize the making of errors, with a de-

CHART 4

CORRELATION CURVE SHOWING EFFECT OF VARYING THE WEIGHT FOR ERRORS


Curve for Group A showing effect on correlation between bonus received and employer's judgment of the employes' abilities in clerical work, when the weighting for errors, used in the figuring of the bonus, is varied from a credit of 10 cents for each error to a deduction of 14 cents for each error.
Coefficients of the zero order.
r_{IA} = .41
r_{IE} = -.30
r_{AE} = .13
σ_A = 369.55
σ_E = 70.98
N = 27
← Present r
Correlation
.55
.50
.45
.40
.35
.30
.25
.20
.15
.10
.05
0
-.05
-.10
Credit for Errors
10 9 8 7 6 5 4 3 2 1
Deduction for Errors in terms of cents
1 2 3 4 5 6 7 8 9 10 11 12 13 14

duction of 4 cents for each error. This would change the correlation between bonus received and the supervisors' ratings from .46 to .54. Further, as the curve shows, the safer course would be to deduct more rather than less than 4 cents, as there is less difference, as reflected in the correlations, between successive possible weights above that amount than below.

This suggests caution against putting too great reliance in the reliability of bonus ratings. And further, it may have a suggestion for business concerns as to a possible method of improving the technique of figuring bonuses so that the bonus will at least reflect the opinions of the officers in charge.

VARIATION IN THE JUDGMENT OF EXPERIENCED EXAMINERS

A short time ago the writer carried out an investigation with the idea of improving the technique of rating papers in composition in civil service examinations for stenographers. In connection with the study there was occasion, as one of the steps in the work, to determine the variation in judgment of various experienced civil service examiners. The results are set forth here as presenting from a little different angle the same factor of variability that has already been emphasized. In the investigation twenty-five compositions were selected from among those submitted in a civil service examination for stenographers. They were mimeographed and sent to some twenty-five civil service examiners, these examiners being asked to rate them according to the directions enclosed—with a full statement as to the type of position concerned and the instructions given candidates in the examination. In Chart 5 on page 36 is seen the variation in judgment of these examiners for each of a number of the compositions. Each dot represents, on the scale, the mark of one examiner. Thus for one of the compositions the range is shown to be from a lowest mark of 40 per cent to one as high as 94 per cent.

A few of the outstanding facts in connection with the marks are noted below.

Some Outstanding Facts in the Rating of the Compositions

In two cases essay 25 received the lowest mark given by an examiner, and in three cases next to the lowest mark; while on the other hand, it was given the highest mark by one examiner, and next to the highest by another.

One examiner gave a passing mark to only six of the compositions, while another examiner gave a passing mark to all but one.

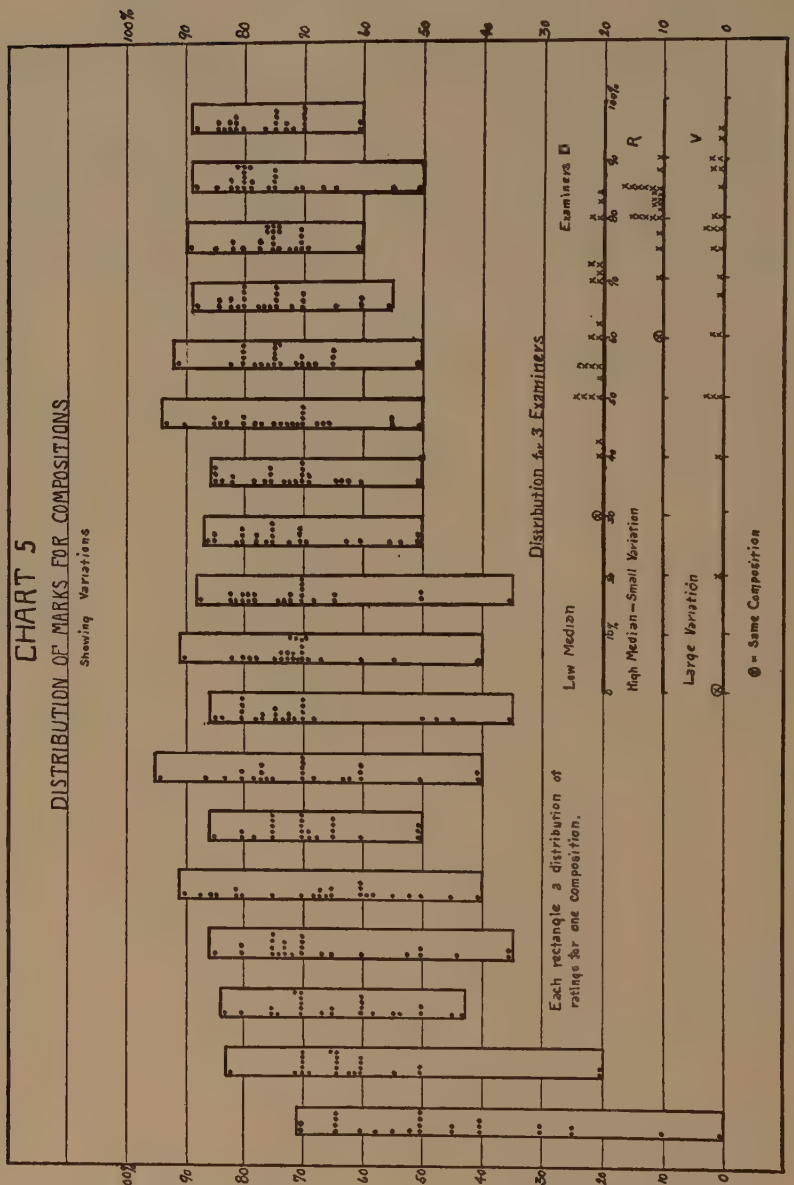


CHART 5. VARIATION IN MARKS OF EXPERIENCED EXAMINERS

One composition was given a mark of zero by one examiner, while another gave it a mark of 70 per cent.

One composition was given a mark of 40 per cent by one examiner, and a mark of 94 per cent by another examiner, the latter considering this composition the best in the set, and the former considering it next to the poorest.

The largest variation, from minimum to maximum in the marks given by any examiner to the set of compositions, is from 0 to 95 per cent, while the smallest is from 65 per cent to 85 per cent.

One examiner gave the same mark to 13 of the 24 compositions, this mark being 70 per cent.

The question arises in connection with all of these facts: Where can we find adequate ratings—adequate and reliable criteria?[1]

CRITERIA STATISTICALLY CONSIDERED

The object of the study just referred to was that of establishing a scale for judging the reliability of raters of English composition. In the study it was assumed that a standard representing the concensus of opinion of a large number of civil service examiners approached more nearly to an ideal for practical application than would the opinion of any one examiner, though that one examiner might be an expert in English. Such was an attempt to statistically work out a criterion.

It must be remembered that in correlations between tests and a criterion, both the criterion and test scores are fallible. In the discussion of this chapter the fallibility of the criterion has been especially emphasized, and various factors that reduce its reliability have been pointed out. When two sets of measures or ratings are obtained—when similar tests are given to the same group under the same conditions, or two or more sets of ratings are secured for the group,—the proper statistical treatment of the results produces data that are significant in judging of the reliability of the measures concerned. Although such use of data has been set forth by others, Kelley[2], in his recently published *Statistical Method*, has discussed the matter in a most helpful way, in its practical applications, and especially where the reliability of criteria is concerned.

[1] This should in no way be construed as reflecting on civil service examinations. Similar results have been found in school marks and would, no doubt, be found in private employment. On the other hand, the fact that civil service commissions are giving attention to these things and facing the facts squarely is truly a most encouraging sign. The commissions are now united in a determination to make their examinations more discriminating, and to use, in tests of fitness, the best that has been advanced by psychology. This concerted effort is represented by the Bureau of Public Personnel Research which is doing real constructive work in this field, the research work of which is directed by Dr. L. L. Thurstone, formerly of Carnegie Institute of Technology.

[2] Kelley, Truman L., *Statistical Method*, Chap. 9. Macmillan Company, 1923.

In connection with his discussion of the estimates of true scores and the probable error of these estimates, Kelley gives an example that might well have been worked out for the judgments of the examiners of the compositions just discussed. The similarity of application, and the indication of the application of statistical methods to practical examination problems should be of especial interest to civil service examiners and men dealing with employment problems. Assuming that proper data have been secured relating to foremen's judgments of the expertness of mechanics, it is not difficult to determine the reliability of such judgments. In the illustration used Kelley gives this as .36. It is then assumed that a trade test is given, and that it shows a correlation of .48 with the criterion, which is the judgment of a single foreman. The judgment of this one foreman is fallible; but we might expect the judgment of a large number to be fallible to a less degree. Kelley assumes that an average of the judgments of an infinite number of foremen would represent a true judgment of expertness, as a mechanician, and upon this basis he shows that the correlation between a single fallible test score and this true criterion would be .80. He says: "If the hiring of a mechanician is not so much for the purpose of satisfying a particular foreman as it is to secure expert workmen, the correlation .80 is not only the one of theoretical importance, but is, in fact, the correct one to use in regression equations estimating expertness from trade test scores."

THE CRITERIA OF THIS STUDY

As to the criteria ratings used in this study, the writer feels safe in saying that, so far as he could determine, they represent the thoughtful and careful judgment of the supervisors giving them. Whatever may have entered into their consideration, it seems quite certain that in the case of the aptitude ratings, the supervisors concerned would have been satisfied, had the clerks been certified to them by the employment department in the order of such ratings. To this extent, then, the criterion may be considered as satisfactory.

An experiment was tried in the case of one supervisor. In the first place, this supervisor gave the ratings for the entire group in rank order, there being forty-nine clerks in the group. About a month later this same supervisor was given two sets of cards, on

those of one set being the names of the clerks to whom she had, in the first case, given even numbered ranks, and on the second set the names of those to whom she had given odd numbered ranks. Thus, each set represented practically the same range of ability, and would average up about the same. The supervisor was asked to arrange the cards (names) of each group, in order of rank and on the same basis as when they were arranged the first time. The supervisor was not aware of the significance of this second rating. In the presence of the writer, without consulting any records, and with quite a remarkable speed, she ranked the cards of each set. The results of this second ranking showed in each case a correlation of .975 with the order of the same clerks in the first rating, when all were rated together. It would seem, therefore, that this supervisor at least had a surety of judgment, whether or not that judgment was in accord with the directions for rating or not. (This ranking was on the basis of actual efficiency in the work.)

APPLICATIONS TO CIVIL SERVICE PROCEDURE

The writer feels that he must not leave this phase of the subject without a suggestion to civil service examiners, especially as this study was stimulated largely by problems arising in connection with the work of a civil service commission.

Practically all civil service examinations are made up of several parts. In an examination for stenographer there may be several different parts including typing, shorthand, spelling, and composition, aside from the rating on experience and training. In making up the composite marks of candidates in such an examination, each one of these parts must be weighted. To put it in terms of the previous discussion—each one of these variables must be weighted. There must be, then, a best weighting. To assign an equal amount to each of the above named variables in a stenographic examination is not in accord with reason. In civil service examinations, then, an attempt is made to assign weights, as nearly as possible, on the basis of the importance of the subject or variable. Naturally such a duty of assigning weights leaves room for a considerable play of individual judgment, and although there may be a general similarity in weighting among different commissions, yet there is enough variation to make a noticeable difference in the final results. The use of the methods of partial and multiple correlation in the study of such a problem should prove helpful and suggestive.

It is possible to determine these best weights for the various parts of a civil service examination by approved statistical procedure.

One fallacy in the present method of weighting parts of an examination may very easily be corrected by the examiner who will give a little attention to the subject, and post himself on the underlying principles of handling statistical data in a comparative way. Let it be assumed that, for the time at least, examiners must proceed on the basis of their best judgments as to the proper weighting to be given to each part of an examination—to each variable. What often happens, in such a case, is that after truly careful and earnest consideration has been given to the assignment of the weights a good part of the value of such effort is lost through the use of the unstatistical method with which the examination data are handled. And yet, as suggested above, this error may be very easily eliminated by an examiner who is willing to give a little study to the problem in order to understand the principles underlying the correction which should be made when weighting such data.

Statistical data (ratings for the different parts of an examination, or for different examiners for the same part) are comparable *only when* proper allowance has been made for the differences in "spread" —differences in scatter—or variations of marks about the averages. In statistical terms this is called "standard deviation," and is expressed as follows $\sigma = \sqrt{\frac{\Sigma (X-M)^2}{N}}$ where M is the average for the group, X stands for an individual gross score or grade, and N is the number of such grades. Σ indicates summation. The standard deviation, then, is the square root of the average of the squares of the deviations of the individual grades from the average grade for the whole group.

The civil service examiner knows that the standards of various markers of examination papers vary considerably, both as to central tendency or average of the set of marks, and as to variation or spread seen among the papers thus marked. Thus in combining sets of marks of two or more examiners, it is not wise to combine their flat percentage marks without giving consideration to their differences in central tendency in marking, and to the spread of the marks given. This same principle also applies when marks for *different subjects* are considered, even though marked by the same examiner. For instance, an examiner would hardly expect to find the same variation or spread in a set of marks for spelling as in one

for composition. In the rating of composition individual judgment has a large chance for play, while in the rating of spelling it has practically none.

The possibilities in this connection are shown in Chart 5, page 36 where are plotted the marks given by three examiners to the same set of 24 composition papers in the experiment described on page 35 of this study. These three sets of marks are certainly not comparable as they stand, in terms of flat percentages, or as would be called in statistical work, "raw scores." Consideration will show that if they were averaged as they stand, the marks of examiner V with his large variation in marks would really carry a much greater weight, when averaged in a combined score, than would the marks of examiner R. It will be seen that in such a combination as that suggested, the less the spread in the marks of an examiner, the less weight they carry, when averaged with those of another examiner with a wide spread in marks. A limiting case of this would appear if the marks of examiner R were all concentrated at one point, say 85 per cent, while the marks of examiner V remained as they are. Then, in averaging, the marks of examiner R would carry absolutely no weight in determining the final order. Their only influence would be that of reducing the range of the marks as thus finally averaged. Thus, the variations of any sets of marks that it is desired to combine are, in themselves, important weighting factors, and unless this is allowed for, the nominal weights, consciously and conscientiously worked out by civil service examiners, are of little avail in the accomplishment of what they have in mind. Exactly this same principle applies when sets of marks on different subjects are combined, whether the marks for each subject were assigned by the same or by different examiners. This then leads to the principle which is stated so effectively by Kelley as follows: "The effective weight given a test is not w, the nominal weight, but $w\sigma$, the product of the nominal weight and the standard deviation of the scores. Accordingly, equally weighted scores are those in which the products of the nominal weights and the standard deviations are equal."

He goes on to say that tests are actually weighted equally only when $w_1\sigma_1 = w_2\sigma_2 = w_3\sigma_3$. Expressed in terms of civil service problems: The marks of each of three examiners, for a given set of papers, are given equal weight in combination only when the standard deviations of each such set of marks are found, and the

sets are then so weighted that the standard deviation of a set, multiplied by this weight, is the same numerically as the product of the standard deviation and weight, for each of the other sets of marks. Thus with sets of marks which it is desired to give equal weight in a combination, it is necessary actually to give such sets of marks *different weights* in order to equalize the weighting effect of their various standard deviations, or spreads. Sometimes the spread or standard deviation of the marks might be very little different, so that the application of this principle would make practically no difference; but the writer believes, from the result of his own experiences in dealing with the marks of civil service examiners, that in a large percentage of cases a material difference would result by applying the above principle, in "averaging" such sets of marks.

To some examiners the standard deviation, with its expression involving squares and square roots, may seem foreign, and have the ring of the theoretical rather than of the practical. A little familiarity, however, with its meaning and use will show its very practical significance, and the ease with which it may be applied. The writer, interested as he is in problems of civil service administration, would strongly urge upon civil service examiners the need of becoming thoroughly familiar with the statistics of the comparability of data, especially as this applies to civil service ratings.

CHAPTER IV

THE SCORING AND WEIGHTING PROBLEM

Section 1

The Scoring Problem, or Weighting for Errors

In scoring a test of this nature, where the time is a constant, and the amount completed in that time varies from individual to individual, the amount done is taken as the basic measure of performance. However, inasmuch as the errors made by different persons also vary they must be given consideration. The question then arises as to the relative value of speed and accuracy in the test, and in the work for which it is selecting.

Every one who has attempted to standardize a test of any kind has met with this problem. The same problem confronts the teachers in our schools. It is present in all types of examinations. In civil service examinations, deductions for errors in typing, for mistakes in transcription of stenographic notes, for misspelled words, and for the numerous other errors, are as a rule determined in a more or less arbitrary fashion. In psychological tests, it has often been considered that a weight of one, for each wrong response, gives about as satisfactory results as any other weighting, and thus it is frequently used.

THE PROBLEM IN A CLERICAL TEST

One who has had experience in an employment office of any kind is familiar with the request for a clerk who is *accurate. She must be accurate!* A slow clerk will be tolerated if she is accurate. Differences in accuracy seem to stand out much more clearly in the mind of many employers than do differences in speed. It is also evident that in actual clerical work carelessness and inaccuracies are looked upon with different degrees of annoyance in different offices. Further, it is apparent that accuracy is a more important factor in determining general office efficiency in one type of work than in another. The relative importance of speed and accuracy depends largely, not only upon the work of the office, but also upon

its organization. In cases where there is an effective system of checking for work, accuracy may be sacrificed to some extent to speed. One thing is quite certain and that is that guessing will result in no great progress in determining the best weighting for errors.

Before entering upon a statistical treatment of the problem, however, one thing should be noted. The tendency to make errors varies according to the form and content of the work or operation. As an illustration: In test 1 the problem of reaction is quite different from what it is in test 5. In test 1 something is being looked for—a difference—and a definite response is made *only when* such a difference is observed. In this test, when attention wanes, the speed of the operation may continue without diminution, but accuracy may be affected to a marked degree. In test 5, on the other hand, a definite reaction is made to every space in the table that the worker comes to, and the error here is in the nature of a definitely wrong response, rather than of a failure to make any response at all. This results in many errors in test 1 and very few in test 5.

ATTACKING THE PROBLEM STATISTICALLY

In connection with this test, the best weighting for errors has been determined by the statistical method. It has been assumed that the best weighting is the one that results in the highest correlation between the test and the criterion. In the beginning of the study, this was undertaken by use of the rank method. For each weight from 0 to 15 the resulting set of scores for a particular test was secured. The correlation of each such set with the criterion was then found, and the best weighting was selected—the one resulting in the highest correlation.

The scoring formula used in this method was $S = A + CE$, where S is the resulting score, A the amount done or items covered in the time, E the errors or wrongs, and C the weight to be applied. If deductions are to be made for errors (rather than credit being given for errors) as of course is the natural expectation, the C will be negative, thus making the formula $S = A - CE$. According to the method just suggested, which is long and tedious, it is not possible to determine the best weight to use without working out the correlation for a considerable series of scores resulting from the application of different weights.

Dr. L. L. Thurstone, formerly of the Carnegie Institute of Tech-

nology and now in charge of the research work for the Bureau of Public Personnel Administration, first suggested the use of the following equation in determining the best weight to use in scoring.[1] Where Thurstone used R for rights and W for wrongs, in this study A is used for amount done and E for errors. Otherwise the equation is the same.[2]

$$C = \frac{\sigma_A\,(r_{IA} \cdot r_{AE} - r_{IE})}{\sigma_E\,(r_{IE} \cdot r_{AE} - r_{IA})} \quad \ldots\ldots\ldots (1)$$

This equation gives the weight C which will result in the highest correlation between the criterion and the scores thus secured, by the use of the scoring formula $S = A + CE$. The weights resulting when this method is used (with the aptitude ratings taken as the criterion) are given in Table II. In the last column of this table is given the average of the different group weights for each test, while the median weight for each such set is indicated by the black face figures.

TABLE II

BEST WEIGHTINGS FOR ERRORS

(APTITUDE CRITERION USED)

Test	Group A	B	C	D	E	F	G	Av.
1	**—3.08**	—1.08	—4.86	—1.55	—3.03	—3.55	*	—2.86
2	—10.12	—6.21	—6.48	—6.59	—12.10	**—9.90**	*	—8.57
3	—2.92	—2.78	—.77	—3.42	—.74	—.46	**—1.03**	—1.73
4	**—2.87**	—10.59	+10.42	—7.25	+5.80	+1.67	—7.24	—1.44
5	+2.59	+1.58	—.05	—23.73	**—4.04**	—9.05	—5.69	—5.48
6	**—1.07**	+7.34	—.83	—6.44	+2.97	—1.24	—19.83	—2.73

* In test 2 for this group, the correlation between the criterion and amount done is zero, and is .0038 between amount done and errors, resulting in the rather astonishing "best weight" of +1286.39, while in test 1 the best weight comes out as 125. In figuring the above averages these two "freak" weights are not used. In the case of such weights it will probably be found in most cases that the curves are practically straight lines, indicating inconsequential differences resulting from varying the weights for errors.

It will be seen from the table that in some cases actual credit for errors (positive weight) results from the use of this equation.

If it were desired to secure a best weighting for errors, to use in a clerical test for clerks in general, it would not be well to use either the average or median weight indicated without first studying the curves as plotted, to find the variation in correlation caused by

[1] Thurstone, L. L., "The Scoring Problem for Mental Tests," *Psychological Bulletin*, Vol. 16, No. 7, pp. 235–40 (July, 1919).

[2] I = Criterion (Aptitude); A = Amount done; E = Errors; C = Weight for Errors; σ = Sigma or Standard Deviation.

changes in weighting for each test and each group concerned. The weights presented in the table have of themselves no great significance. To be of value they must be supplemented by a curve, or set of correlations, showing the effect of variation in weighting. The significance of this statement will be seen from a study of the curves shown in Charts 6 to 8. These curves were plotted from the data resulting from the substitution in the following equation of the appropriate correlation coefficients; of standard deviations (which in the equation were treated as constants); and of varying values of C from — 10 to + 10

$$r_{1(A+CE)} = \frac{r_{IA}\cdot\sigma_A + C\cdot r_{IE}\cdot\sigma_E}{\sqrt{\sigma_A^2 + 2\cdot C\cdot r_{AE}\cdot\sigma_A\cdot\sigma_E + C^2\cdot\sigma_E^2}} \quad \ldots\ldots\ldots (2)$$

A STUDY OF THE WEIGHTING CURVES

A study of the curves in connection with the data given in Table II will serve to bring out the truth of the statement made, as to the need of a careful consideration of the curves. From the table it is found that the best weighting for test 5, as applied to group D, is —23.73. The best weighting indicated in the table, for the same test, but for group B is +1.58. Suppose it were desired to find the best weighting for this test as it might be used in the selection of a general list of clerks to be certified for work in either of these groups (D or B). If the average of these two weights were taken it would be —11.15. When the curves of Charts 6 and 7 are examined, the effect of this averaging upon the correlation is easily seen. The *slope* of the curve for test 5, group E (page 47), is very little, and it makes practically no difference in the resulting correlation whether no deduction is made for errors, or whether the deduction is that indicated in the table as the best: —23.73. In both cases the correlation is .46, all the variation being in the third decimal place. Thus, using the average weight suggested (—11.15) would not change the correlation for this test any appreciable amount; while for group B where the slope of the curve for this test is quite steep, a change of the weighting from the 1.58 to —11.15 would reduce the correlation for that group from .31 to .12. Further, by keeping the weight at 1.58—the best weight for group B—the correlation would not be appreciably reduced for either group. In such a case, had the investigator gone ahead, using averages without studying the curves suggested, he would have unnecessarily lowered the value of the test to a considerable extent.

CHART 6

CORRELATION CURVES FOR GROUP E

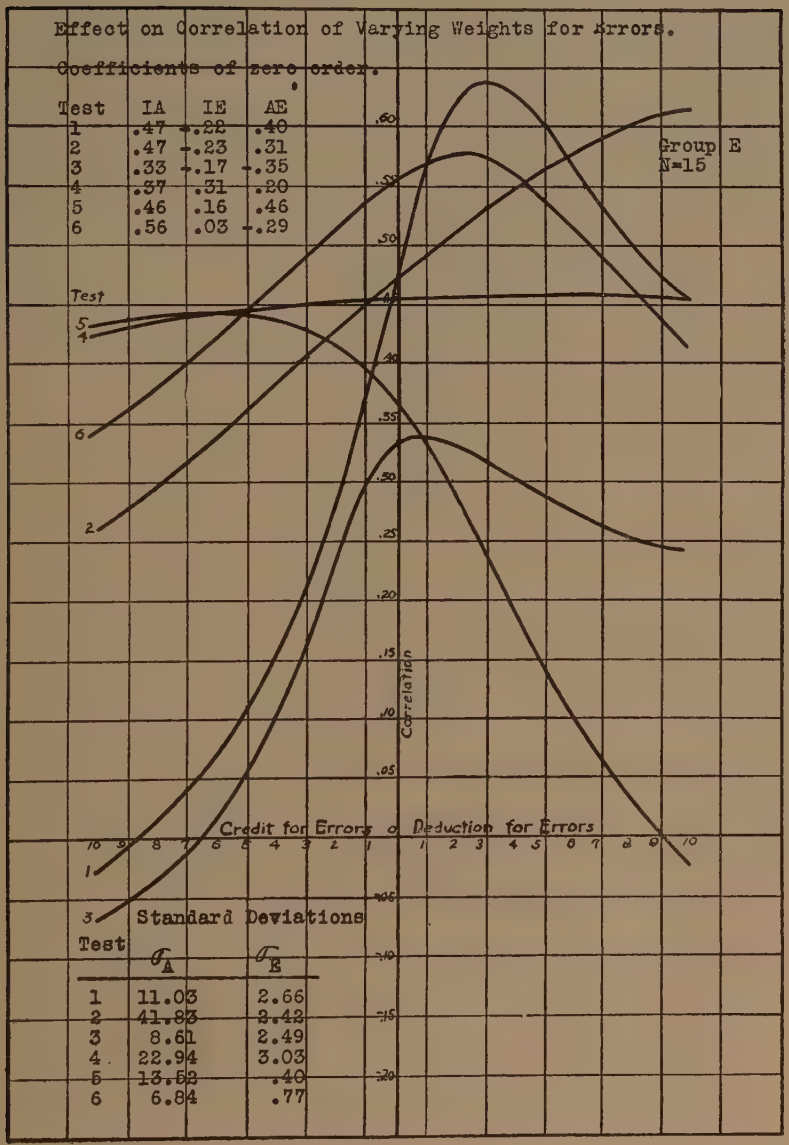


CHART 7

CORRELATION CURVES FOR GROUP B

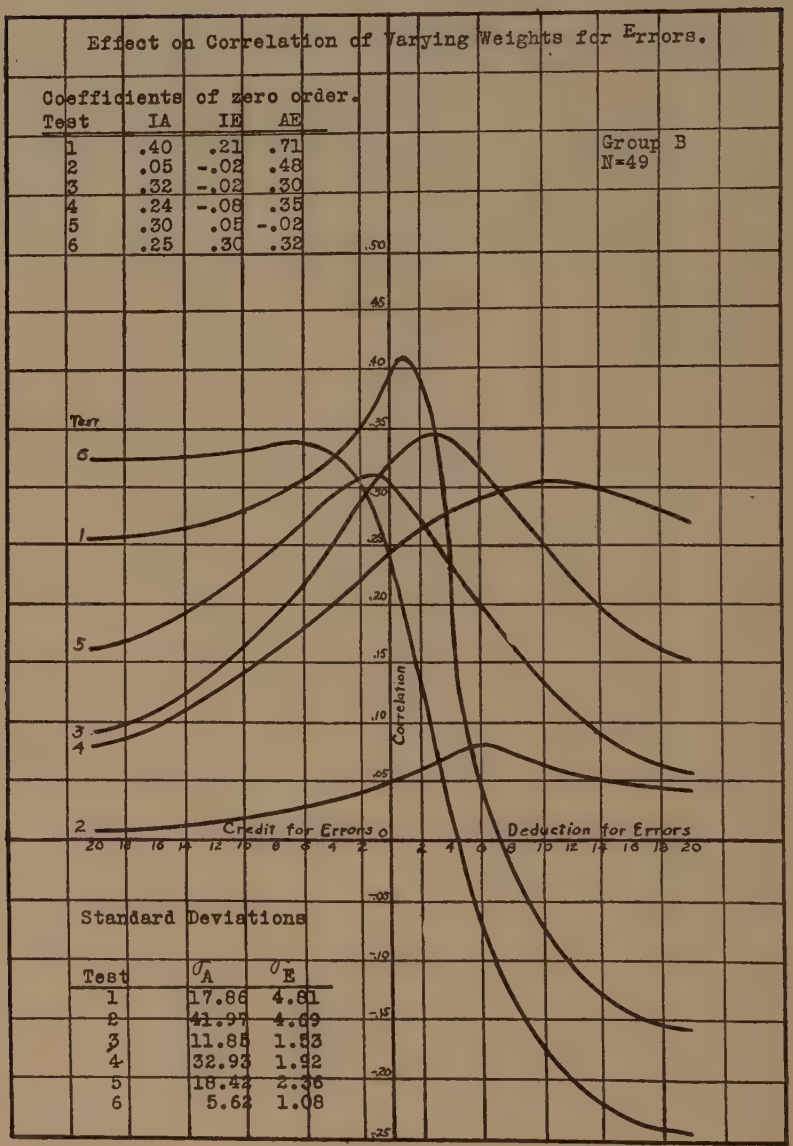
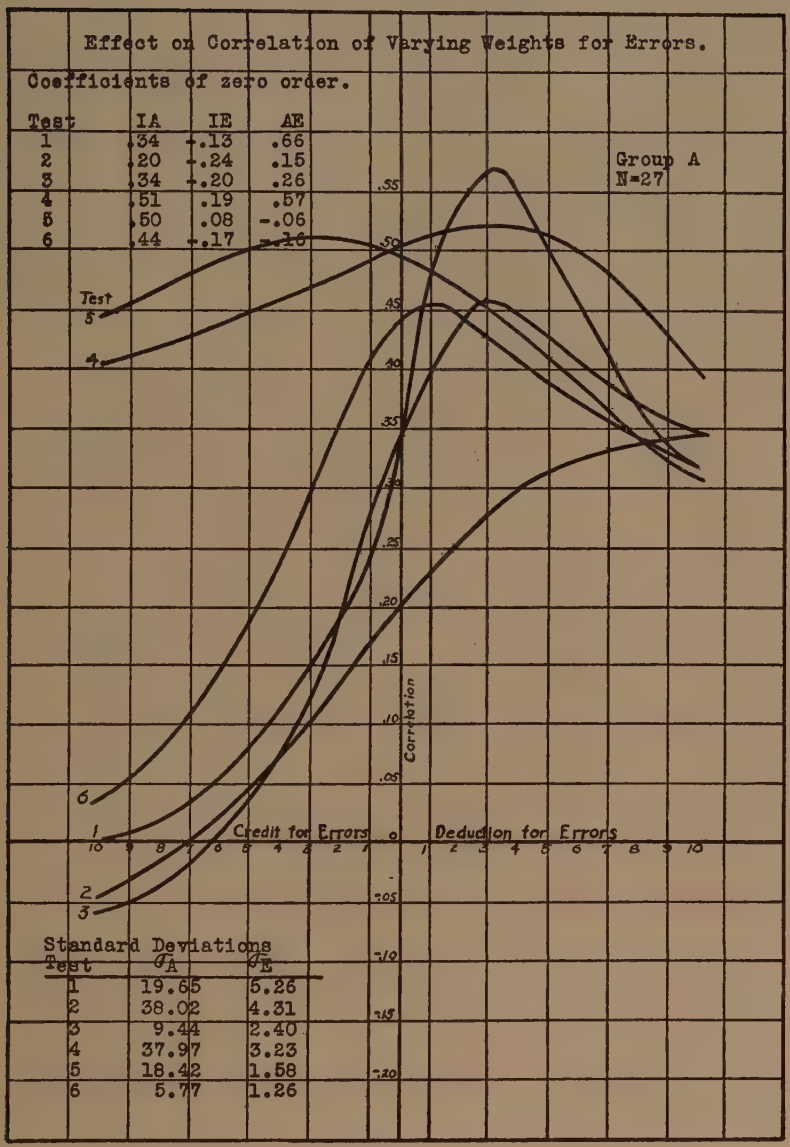


CHART 8

CORRELATION CURVES FOR GROUP A



A study of the curves thus indicates the significance of this variation of weighting for errors. When the curve is flat, variations count for very little, while with a steep slope a little variation may make a marked difference. In the three charts, Charts 6, 7, and 8, the curves for test 1 show a very steep slope.

The three sets of curves are given to enable the reader to see the fluctuations as well as the similarities in the curves as they appear from group to group. (The deductions or negative weights are given on the right of the vertical axis rather than following the usual order of having negative values at the left.)

These curves might be subjected to a detailed analysis and study which, if carried very far, would lead into the realms of higher mathematics; but that is not within the province of this discussion. However, a study of the curves, along the lines suggested, should serve to indicate the effect of weighting for errors upon the size of correlations, as well as caution against the dangers of proceeding to make use of the results of equations without a sufficient consideration of the underlying facts and conditions. In connection with each diagram are given the constants used in the equation, for the benefit of such readers as care to study the curves more in detail.

A TESTING FOR CONSISTENCY IN THE CURVES

Some of the curves seem to have similarities as they appear in connection with the three different groups. The question is thus raised as to what extent one might expect a similarity of curve, and thus of weighting, from group to group, for the same test. In other words, suppose that a best weighting for test 2 were determined for a group of a hundred persons who had taken a civil service examination, to what extent might such a weighting be expected to hold good for other examinations where the same test was used but where the groups were different, but supposedly of the same general make-up as the trial group? An experiment was attempted with the hope of getting some suggestions on this point.

Previous to this particular experiment five groups of 100 each had been selected from among those candidates who took this test as a part of a state civil service examination for the position of "Clerk." Candidates had been assigned, by the writer, to one or another of five groups on the basis of drawings from a box of numbered cards thoroughly mixed. The groups were therefore

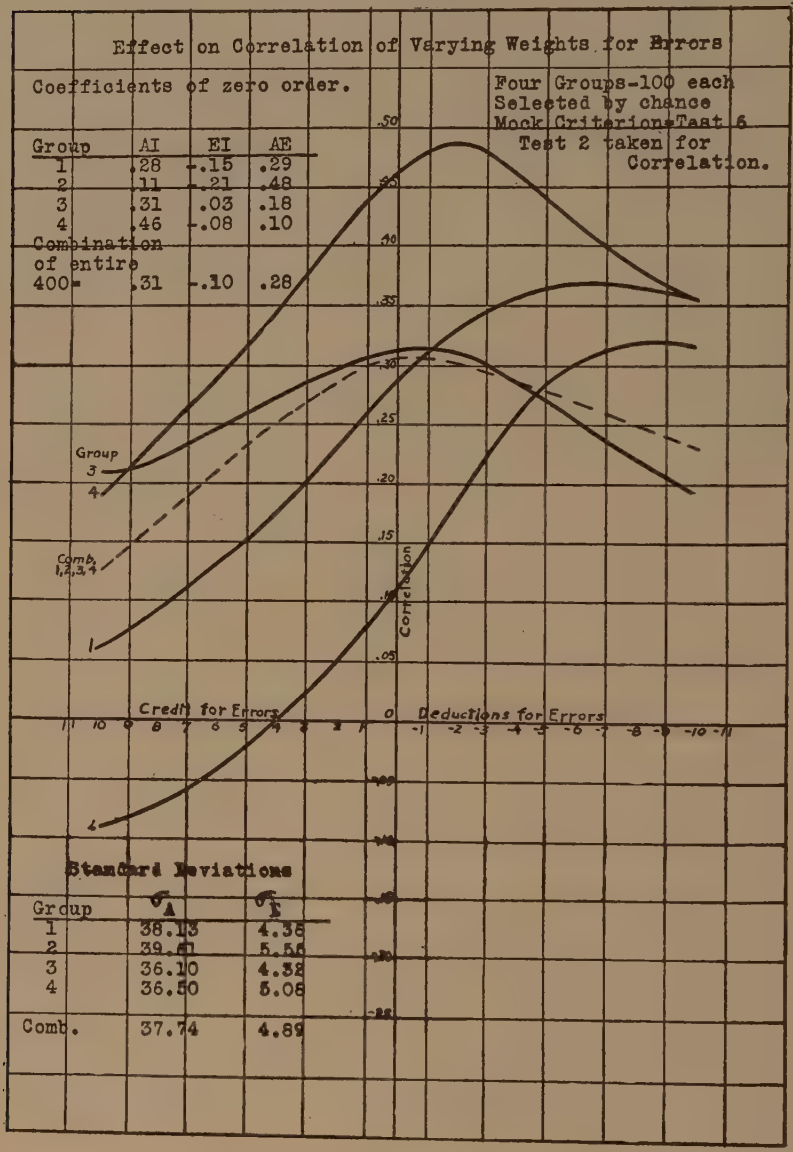
made up so far as possible on the basis of chance. From among these five groups four were selected (by chance) for this experiment.

Inasmuch as there was no aptitude or efficiency criterion available, some other variable had to be selected as a sort of mock substitute in this experiment. As the best available substitute of this kind, test 6 was selected, an arbitrary weight of 3 given to errors and the score for the test computed accordingly. For the purpose desired, this was considered to serve very well as a substitute for a regular criterion. It was objective in nature and determined by work done (in an arithmetic test) with a constant penalty for errors. This would thus correspond to a certain extent with a supervisor's judgment of a clerk's ability, influenced as it must be by annoyance with errors. Test 2 was then taken for experiment in determining the curves of correlation with this "fake" criterion—ratings on test 6. The correlations resulting from varying the weights for errors in test 2 were then worked out and plotted. Equation 2 on page 46 was used. The resulting curves for each of the four groups are given in Chart 9 on page 52. The dotted line shows the curve resulting when the data of all four groups are combined, and then the correlations found.

The maximum correlation values of the four curves, as well as the best weightings, are seen to vary to a considerable degree. This naturally gives rise to the question of the use of bothering about weights if their results are so uncertain. But it must be remembered that *the weighting of the errors of a test cannot be escaped.* When no weighting at all is attempted it simply means that a "zero" weight is being given, even though by neglect rather than consciously. Further, if the problem is settled by giving a weight of one, and "being done with it," the skeptical attitude should hold just as truly as though a larger weight had been contemplated, unless there is evidence to show that such a weight is subject to less variation from curve to curve. It will be noted that the correlation of "amount done" with this mock criterion—the correlation with zero deductions for errors for test 2—varies from .12 to .46, and this with groups supposed to be similar, so far as chance could direct. An experiment with a different test might have presented quite different results, and the selection of this mock criterion may not have been wise; but back of all is the fact of the marked variations in the correlations, making up the equation from which the curves were plotted.

CHART 9

COMPARISON OF CORRELATION CURVES FOR FOUR SIMILAR GROUPS



This phase of the discussion must here be left with a closing suggestion of the need for a more intensive study of the subject in all its aspects and a further accumulation of data.

Section 2

The Problem of Selecting and Weighting Tests for a Battery

The next problem is that of determining the relative effectiveness of the tests, as well as the best weighting for such tests as are selected. The usual procedure in test standardization has been to determine upon a deduction for errors, in the scoring of the individual tests, and then to find the best weighting for the scored tests, as they are used in the battery. Those familiar with the statistics of correlation know that, although the task of finding the multiple correlation for a group of tests is not an unreasonably lengthy process, the determination of regression equations by which the best weighting for each variable is found, is a very long and tedious process. This is true even when a very few variables are considered.

Dr. Herbert A. Toops has derived a method which enables one, with very much less work, to take a group of tests and by selection of the "key" or basic test—the one that shows the highest correlation as it stands alone—to add to it, one by one, from among the other tests in such a way that each successive test added is the one which, at that point in the procedure, has the greatest effect in raising the correlation. The method also shows, at each point, just what the maximum or "multiple ratio" correlation is, and the weight that should be assigned to each test in securing such correlation. In other words, the regression equation is built up as the process advances. The "multiple ratio" correlation of this method, although not exactly the same as the ordinary multiple correlation, approximates it very closely.

One can proceed, step by step, to the point where the addition of another test does not add sufficiently to the correlation to warrant the extra work that would be necessary if it, or the remaining tests, were included in the battery being arranged. In the application of the method in what follows, the variables of a test ("amount done" and "errors") were not added when they contributed less than .01 to the multiple ratio correlation.

RESULTING REGRESSION EQUATIONS.

In this study the method was used with the data for several groups of clerks, without previously making deductions for errors—simply dealing with each set of "errors" as a variable, as though the error scores were being considered as results of a separate test instead of only one of the two variables of the same test.[1] Equation 3 which follows is the regression equation resulting from the application of this method to the variables of the tests for group A.

$$\frac{x_s}{\sigma_s} = 1.00\frac{x_{4a}}{\sigma_{4a}} - .24\frac{x_{4e}}{\sigma_{4e}} + .81\frac{x_{5a}}{\sigma_{5a}} + .03\ \frac{x_{5e}}{\sigma_{5e}} - .36\frac{x_{6a}}{\sigma_{6a}} - .04\frac{x_{6e}}{\sigma_{6e}} \quad . \ . \ . \quad (3)$$

In this equation the subscripts designate the variables: a for the "amount" variable, and e for the "error" variable—for the tests indicated by the numbers. Thus x_{4a} refers to the "amount" variable for test 4; x_s stands for the regression value in terms of deviation.

The weights or coefficients in the above equation are taken directly from the correlation chart used in this method, and are what are referred to by Toops as the beta (β) weights. These weights cannot be applied directly to the gross scores, or actual scores of the test, but only to scores expressed as deviations from the mean in terms of standard deviations, as in equation 3. If, then, it is desired to have the equation such that the coefficients or weights may be applied directly to gross scores, the deviations of equation 3 must be translated into terms of gross scores, and the resulting form of the weights or coefficients noted.

The gross weight (W) expressed in terms of the beta weight (β) and sigma (σ) or standard deviation is as follows:

$$W = \frac{\beta}{\sigma}$$

Further, deviations may be expressed in terms of gross scores and means, as follows:

$$x = X - M$$

where x stands for scores in terms of deviation, X for gross scores and M for the mean of the set of scores under consideration. If therefore the expression $X - M$ is substituted for x or deviations in equation 3 and both members of the equation multiplied through

[1] For each test any examinee originally had two scores: "amount done"—in terms of number of items covered in the time allowed—and "errors" or number of mistakes made.

by σ_s, we have equation 4, where X_s is the regression value in terms of gross scores.

$$X_s - M_s = \frac{\sigma_s}{\sigma_{4a}}(X_{4a} - M_{4a}) - \sigma_s \frac{.24}{\sigma_{4e}}(X_{4e} - M_{4e}) + \sigma_s \frac{.81}{\sigma_{5a}}$$
$$(X_{5a} - M_{5a}) \text{ etc.} \ldots\ldots\ldots \quad (4)$$

Since σ_s appears in every term on the right side of the equation, its effect on the correlations of S and criterion scores is nil and it may be disregarded. If now all the Ms or mean values are combined there results a constant K as shown in equation 5, which constant may be neglected in the practical use of the equation, for the same reason—that the correlation between S and I is uninfluenced by such omission. In his use of this method[1], Toops speaks of this constant as "an arbitrary constant to be taken of such magnitude that the criterion scores will be readily handled quantities." It may be given a value of zero.

$$X_s = \frac{1}{\sigma_{4a}} X_{4a} - \frac{.24}{\sigma_{4e}} X_{4e} + \frac{.81}{\sigma_{5a}} X_{5a} + \frac{.03}{\sigma_{5e}} X_{5e} + \frac{.36}{\sigma_{6a}} X_{6a} -$$
$$\frac{.04}{\sigma_{6e}} X_{6e} + K \ldots\ldots\ldots \quad (5)$$

When the beta weights of equation 5 (1, .24, .81, .03, etc.) are actually divided by the corresponding sigmas as indicated, the result is equation 6, expressed in terms of gross scores, the coefficients being in terms of gross score weights. With these gross score weights used as multipliers in connection with the respective gross scores of the tests, the three tests thus combined function in the best possible manner as a selective instrument for group A, on the basis of the method used.

$$X_s = .026X_{4a} - .076X_{4e} + .044X_{5a} + .019X_{5e} + .062X_{6a} -$$
$$.032X_{6e} \ldots\ldots\ldots \quad (6)$$

Equations 7 and 8 are regression equations for the selection and weighting of tests for groups B and E, given in the form of equation 6. Above each variable, however, is shown in parentheses the "multiple ratio" correlation resulting from the addition of that variable to the battery being constructed. Thus may be seen the amount contributed by each additional variable. The variables appear in the equations in the same order in which they were se-

[1] For full discussion of Toop's method, and complete formulae, see dissertation by Miss Evelyn Garfiel, on *Tests of Motor Ability*, Columbia University, Archives of Psychology, No. 62.

lected by the Toops' method. As already stated, however, variables of "amount" are in every case added before the corresponding "error" variable, and thus, as is apparent in some cases, the "error" variable would, had it been considered entirely independently of the "amount" variable, have been added before it actually was, because of its larger contribution to the battery. In other words the errors of a test, in such cases, contribute more to the predictive total than does "amount done," but this error variable is not added until after its corresponding "amount" variable is taken into the battery.

REGRESSION EQUATIONS

GROUP B

$$X_s = \overset{(.400)}{.056X_{1a}} - \overset{(.413)}{.060X_{1e}} + \overset{(.458)}{.042X_{3a}} - \overset{(.473)}{.177X_{3e}} - \overset{(.483)}{.005X_{2a}} + \overset{(.483)}{.001X_{2e}} \quad \ldots\ldots\ldots (7)$$

GROUP E

$$X_s = \overset{(.557)}{.146X_{6a}} + \overset{(.591)}{.429X_{6e}} + \overset{(.635)}{.012X_{2a}} - \overset{(.677)}{.194X_{2e}} + \overset{(.686)}{.027X_{1a}} - \overset{(.780)}{.289X_{1e}} - \overset{(.819)}{.062X_{3a}} - \overset{(.840)}{.116X_{3e}} - \overset{(.840)}{.005X_{5a}} - \overset{(.847)}{.400X_{5e}} \quad \ldots\ldots\ldots (8)$$

The multiple ratio correlation that would have resulted, had the entire number of tests been included in the battery, may easily be estimated when it is remembered that according to the method used none of the remaining tests would have added .01 to the correlation. For group E, test 5 would not be used because, as shown in equation 8, 5a and 5e together raise the correlation only from .840 to .847—less than .01.

Inasmuch as the chief interest of this study was to consider the individual values of the different tests, comparatively little use was made of the multiple ratio correlation. Further, in the use of this method the small number of measures in each group would keep the resulting correlations from being of any great significance.

CHAPTER V

RESULTING CORRELATIONS AND INTERPRETATIONS

METHODS AND RESULTING DATA

In this study, three correlation methods or devices were used, all, however, based on the Pearson product-moment method. In the work of finding the correlations of the first order—the straight correlation between *two* variables—Toops' short-cut method was followed.[1] In finding the multiple correlation (R) in connection with each test and each group, as determined by the most effective weighting for errors, the usual formula, suggested by Thurstone in his article on "A Scoring Method for Mental Tests,"[2] was used. The criterion scores were of course first changed into terms of their standard deviations, on the assumption of normal distribution. As has also been stated, the aptitude criterion was used in determining the R for the combinations of tests.

On pages 58 to 60 are several tables giving the significant and fundamental data resulting from this study. It seemed that the data in this form, and given in detail, would be of the greatest value to those working in this field, and interested in making their own interpretations from the more fundamental and underlying data. In connection with each intercorrelation is given the correlation resulting from a combination of all the groups into one, not through averaging but from a combination of original scores. This will perhaps prove of interest to those who care to examine these data from a statistical standpoint.

THE POSSIBLE USE OF THE TESTS IN THE SELECTION OF GENERAL CLERKS

In civil service work the call is most frequently for "general clerks"—those who can be placed at any type of simple clerical work, where they may have to perform any of a number of various types of duties. In that case, the discrimination used in selecting particular tests for particular types of work cannot be used.

[1] Toops, Herbert A., "Computing Intercorrelations on the Adding Machine," *Journal of Applied Psychology*, Vol. VI, pp. 172–184 (June, 1922).

[2] Thurstone, L. L., "A Scoring Method for Mental Tests," *Psychological Bulletin*, Vol. XVI, No. 7, pp. 235–40 (July, 1919).

GENERAL TABLE III

Intercorrelations for All Variables and Groups

Groups		Test 1 am't	Test 1 e errors	Test 2	Test 2 e	Test 3	Test 3 e	Test 4	Test 4 e	Test 5	Test 5 e	Test 6	Test 6 e	Max. Cor.[1] *RI*	Max. Cor.[1] *RI'*
Varia-	A		.66	—.15	—.39	.32	.28	.24	—.11	.53	.07	.31	.13	.34	.22
bles	B		.71	.36	.24	.31	.09	.32	—.06	.47	—.05	.40	.14	.40	—.01
	C		.25	.65	.60	.41	.34	.40	.13	.57	—.09	.50	—.32	.38	.16
Test	D		.60	.52	.06	.41	.29	.19	.26	.40	—.10	.46	—.27	.51	.28
1	E		.40	.32	—.17	.26	.40	.54	.26	.66	.39	.64	—.35	.47	—.07
(am't)	F		.77	.22	—.10	.62	.08	.33	—.02	.70	—.12	.07	.07	.17	.21
	G		.55	.51	.10	.23	.06	.24	—.26	.44	.03	.16	—.25	—.04	.21
	H		.60	.69	—.19	.49	.26	.42	—.47	.30	.13	.68	—.46		.24
Comb.[2]			.64	.40	.09	.31	.22	.38	.06	.50	.04	.42	—.01		
	A	.66		—.34	—.29	—.16	.41	—.22	—.15	.15	.19	—.09	.32	—.13	.11
	B	.71		.17	.31	.14	.06	.06	—.08	.13	.01	—.01	.11	.21	—.01
	C	.25		—.21	.14	—.02	—.08	—.23	—.10	—.15	.00	—.19	.09	—.09	—.25
	D	.60		.20	.29	—.18	.12	—.20	.21	.11	—.20	—.02	.09	.11	—.12
e	E	.40		.22	—.18	—.21	.26	.08	.21	.04	.13	.06	—.16	—.22	—.41
(errors)	F	.77		.04	—.07	.26	—.01	.15	—.12	.33	.10	—.20	—.09	.05	.09
	G	.55		.30	—.30	—.23	—.18	—.41	—.47	—.13	.51	.09	.26	—.06	.10
	H	.60		.46	.19	.09	.39	.67	—.06	.29	—.10	.20	—.04		.04
Comb.		.64		.07	.17	—.04	.16	.06	.02	.14	.07	.04	.15		
	A	—.15	—.34		.15	.54	.04	.24	.31	.22	—.24	.11	—.18	.20	.05
	B	.36	.17		.48	.08	.08	.31	.14	.22	.01	.23	—.08	.05	.06
	C	.65	—.21		.46	.40	.26	.52	.23	.83	—.11	.73	—.42	.72	.50
	D	.52	.20		.52	.17	—.03	.01	—.07	.26	.06	.48	.00	.38	.16
2	E	.32	.22		.31	.07	.08	.31	.45	.25	.34	.40	.09	.47	.21
	F	.22	.04		.03	.36	—.20	.04	—.24	.39	—.07	.33	—.30	.33	.43
	G	.51	.30		—.00	.29	.22	.21	—.17	.63	—.05	—.10	.08	.00	.21
	H	.69	.46		—.14	.52	.00	.41	—.54	.45	.12	.72	—.39		.40
Comb.		.40	.07		.22	.24	.12	.41	.17	.41	—.01	.45	—.16		
	A	—.39	—.29	.15		.04	.17	—.09	—.21	—.44	—.14	—.48	.04	—.24	—.23
	B	.24	.31	.48		—.22	—.04	—.02	.21	—.04	.04	—.11	—.13	—.02	.17
	C	.60	.14	.46		.17	.20	.21	.26	.45	—.15	.45	—.18	.19	.09
	D	.06	.29	.52		—.09	—.03	—.13	—.05	.02	.08	—.11	.35	—.18	—.45
e	E	—.17	—.18	.31		—.25	.36	.05	—.02	—.32	.10	—.30	.41	—.23	.17
	F	—.10	—.07	.03		—.08	.24	.18	.28	—.26	.41	—.20	.45	—.45	—.50
	G	.10	—.30	.00		.02	.33	.36	.50	.36	—.18	—.48	—.34	—.18	.05
	H	—.19	.19	—.14		—.16	.42	.39	.62	.02	.22	—.43	.65		—.49
Comb.		.09	.17	.22		—.07	.12	.02	.10	—.03	.06	—.13	.23		
	A	.32	—.16	.54	.04		.26	.26	—.21	.51	—.16	.27	—.17	.34	—.01
	B	.31	.14	.08	—.22		.30	.44	—.25	.45	.04	.30	.19	.32	.02
	C	.41	—.02	.40	.17		.51	.56	.33	.40	.29	.13	—.20	.30	—.09
	D	.41	—.18	.17	—.09		.58	.56	.29	.48	.12	.46	—.36	.36	.21
3	E	.26	—.21	.07	—.25		—.35	.11	.25	.59	.01	.72	—.17	.33	.11
	F	.62	.26	.36	—.08		.33	.26	.01	.67	—.41	.40	.23	.52	.58
	G	.23	—.23	.29	.02		.12	.19	—.19	.59	—.05	.24	—.07	.33	.11
	H	.49	.09	.52	—.16		.08	.40	—.55	.62	.30	.71	—.47		.29
Comb.		.31	—.04	.24	—.07		.25	.32	—.06	.52	.04	.26	—.08		
	A	.28	.41	.04	.17	.26		—.12	—.17	—.21	.23	—.14	.58	—.20	—.09
	B	.09	.06	.08	—.04	.30		—.01	—.04	—.04	—.08	.06	.24	—.02	.05
	C	.34	—.08	.26	.20	.51		.20	.52	.27	—.15	.11	—.23	.12	.04
	D	.29	.12	—.03	—.03	.58		.23	.23	.29	.21	.08	—.25	.00	—.10
e	E	.40	.26	.08	.36	—.35		.21	.33	.03	.25	.07	—.07	—.17	.00
	F	.08	—.01	—.20	.24	.33		.40	.37	.27	—.18	.21	.72	.09	.19
	G	.06	—.18	.22	.33	.12		.28	—.11	.12	—.37	.08	—.18	—.02	.14
	H	.26	.39	.00	.42	.08		.62	.32	.35	.52	—.07	.21		—.66
Comb.		.22	.16	.12	.12	.25		.30	.32	.16	.06	.13	.31		

[1] In the last two columns are the maximum correlations of each variable (for each group) with the aptitude criterion—*RI*; and with the efficiency criterion—*RI'*. (See pages 28 and 29.)

[2] For explanation of this combination score see ¶ 2 on page 57.

GENERAL TABLE III (Continued)

		Variables												Max. Cor.	
	Groups	Test 1 am't	e errors	Test 2	e	Test 3	e	Test 4	e	Test 5	e	Test 6	e	*RI*	*RI'*
	A	.24	—.22	.24	—.09	.26	—.12		.57	.34	.25	.56	—.16	.50	.31
	B	.32	.06	.31	—.02	.44	—.01		.35	.42	—.12	.52	.13	.24	—.10
	C	.40	—.23	.52	.21	.56	.20		.41	.27	.25	.42	—.22	.35	—.13
	D	.19	—.20	.01	—.13	.56	.23		.50	.40	—.14	.35	—.08	.22	—.01
4	E	.54	.08	.31	.05	.11	.21		.20	.15	.49	.36	.22	.37	—.03
	F	.33	.15	.04	.18	.26	.40		.60	.45	—.05	.43	.20	—.20	—.02
	G	.24	—.41	.21	.36	.19	.28		.42	.44	—.39	.21	—.27	—.06	.12
	H	.42	.67	.41	.39	.40	.62		—.05	.63	.39	.31	.19		—.30
	Comb.	.38	.06	.41	.02	.32	.30		.51	.48	.04	.53	.06		
	A	—.11	—.15	.31	—.21	—.21	—.17	.57		.14	.18	.31	.04	.19	.05
	B	—.06	—.08	.14	.21	—.25	—.04	.35		—.04	—.10	.11	.19	—.08	—.09
	C	.13	—.10	.23	.26	.33	.52	.41		.23	—.03	.35	—.19	.34	.25
	D	.26	.21	—.07	—.05	.29	.23	.50		.00	—.12	.22	.24	—.13	—.18
e	E	.26	.21	.45	—.02	.25	.33	.20		.17	.17	.53	.24	.31	.30
	F	—.02	—.12	—.24	.28	.01	.37	.60		—.04	.04	.36	.29	—.15	—.13
	G	—.26	—.47	—.17	.50	—.19	—.11	.42		—.05	.00	—.26	—.27	.01	—.01
	H	—.47	—.06	—.54	.62	—.55	.32	—.05		—.32	—.25	—.66	.61		—.54
	Comb.	.06	.02	.17	.10	—.06	.32	.51		.13	—.02	.26	.27		
	A	.53	.15	.22	—.44	.51	—.21	.34	.14		—.06	.35	—.22	.50	.11
	B	.47	.13	.22	—.04	.45	—.04	.42	—.04		—.02	.42	.08	.30	—.06
	C	.57	—.15	.83	.45	.40	.27	.27	.23		—.09	.59	—.33	.56	.55
	D	.40	.11	.26	.02	.48	.29	.40	.00		.19	.47	—.03	.12	—.06
5	E	.66	.04	.25	—.32	.59	.03	.15	.17		.46	.80	—.47	.46	.14
	F	.70	.33	.39	—.26	.67	.27	.45	—.04		—.54	.49	.18	.35	.47
	G	.44	—.13	.63	.36	.59	.12	.44	—.05		—.35	—.21	—.17	.29	.39
	H	.30	.29	.45	.02	.62	.35	.63	—.32		.32	.57	—.26		—.03
	Comb.	.50	.14	.41	—.03	.52	.16	.48	.13		.04	.40	—.09		
	A	.07	.19	—.24	—.14	.16	.23	.25	.18	—.06		.07	.46	.08	.36
	B	—.05	.01	.01	.04	.04	—.08	—.12	—.10	—.02		.03	.24	.05	.27
	C	—.09	.00	—.11	—.15	.29	—.15	.25	—.03	—.09		—.14	.34	—.06	—.18
	D	—.10	—.20	.06	.08	.12	.21	—.14	—.12	.19		.06	—.02	—.19	—.22
e	E	.39	.13	.34	.10	.01	.25	.49	.17	.46		.37	—.04	.16	.11
	F	—.12	.10	—.07	.41	—.41	—.18	—.05	.04	—.54		—.45	—.07	—.29	—.47
	G	.03	.51	—.05	—.18	—.05	—.37	—.39	.00	—.35		.23	.37	—.15	—.15
	H	.13	—.10	.12	.22	.30	.52	.39	—.25	.32		.00	.05		—.29
	Comb.	.04	.07	—.01	.06	.04	.06	.04	—.02	.04		.01	.16		
	A	.31	—.09	.11	—.48	.27	—.14	.56	.31	.35	.07		—.16	.44	.26
	B	.40	—.01	.23	—.11	.30	.06	.52	.11	.42	.03		.32	.25	.04
	C	.50	—.19	.73	.45	.13	.11	.42	.35	.59	—.14		—.10	.65	.57
	D	.46	—.02	.48	—.11	.46	.08	.35	.22	.47	.06		—.11	.48	.41
6	E	.64	.06	.40	—.30	.72	.07	.36	.53	.80	.37		—.29	.56	.28
	F	.07	—.20	.33	—.20	.40	.21	.43	.36	.49	—.45		.11	.17	.36
	G	.16	.09	—.10	—.48	.24	.08	.21	—.26	—.21	.23		—.18	.10	.05
	H	.68	.20	.72	—.43	.71	—.07	.31	—.66	.57	.00		—.49		.36
	Comb.	.42	.04	.45	—.13	.26	.13	.53	.26	.40	.01		.02		
	A	.13	.32	—.18	.04	—.17	.58	—.16	.04	—.22	.46	—.16		—.17	—.03
	B	.14	.11	—.08	—.13	.19	.24	.13	.19	.08	.24	.32		.30	.08
	C	—.32	.09	—.42	—.18	—.20	—.23	—.22	—.19	—.33	.34	—.10		—.14	—.04
	D	—.27	.09	.00	.35	—.36	—.25	—.08	.24	—.03	—.02	—.11		—.56	—.53
e	E	—.35	—.16	.09	.41	—.17	—.07	.22	.24	—.47	—.04	—.29		.03	.34
	F	.07	—.09	—.30	.45	.23	.72	.20	.29	.18	—.07	.11		—.04	.02
	G	—.25	.26	.08	—.34	—.07	—.18	—.27	—.27	—.17	.37	—.18		—.26	—.36
	H	—.46	—.04	—.39	.65	—.47	.21	.19	.61	—.26	.05	—.49			—.36
	Comb.	—.01	.15	—.16	.23	—.08	.31	.06	.27	—.09	.16	.02			

TABLE IIIA

MAXIMUM CORRELATIONS WITH APTITUDE (R_I) AND EFFICIENCY RATINGS ($R_{I'}$)

Group	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6	
	R_I	$R_{I'}$	R_I	$R_{I'}$	R_I	$R_{I'}$	R_I	$R_{I'}$	R_I	$R_{I'}$	R_I	$R_{I'}$
A	.58	.22	.34	.24	.45	.09	.52	.34	.51	.38	.45	.26
B	.41	.01	.07	.18	.35	.05	.30	.12	.31	.28	.34	.08
C	.42	.34	.74	.52	.30	.13	.41	.36	.56	.56	.65	.57
D	.56	.45	.58	.64	.44	.35	.36	.20	.25	.22	.70	.64
E	.65	.42	.62	.23	.34	.11	.44	.31	.46	.15	.59	.52
F	.22	.24	.57	.67	.53	.58	.20	.15	.37	.54	.17	.36
G	.06	.21	.18	.22	.33	.17	.08	.14	.29	.39	.32	.36
Av.	.41		.44		.39		.33		.39		.46	

TABLE IIIB

AMOUNT DONE AND PERCENTAGE OF ERRORS FOR EACH TEST AND GROUP

Grp	Test 1		Test 2		Test 3		Test 4		Test 5		Test 6	
	a	%e	a	%e	a	%e	a	%e	a	%e	a	%e
A	57.3	.125	80.7	.061	16.6	.136	82.5	.039	50.6	.014	16.8	.095
B	52.0	.100	74.0	.093	27.0	.063	62.7	.027	54.3	.014	17.0	.081
C	40.2	.058	92.9	.025	16.6	.097	70.2	.041	38.0	.004	18.2	.029
D	59.0	.113	83.4	.083	11.2	.156	77.0	.053	50.8	.010	22.7	.068
E	52.8	.056	113.0	.025	27.0	.101	71.9	.046	59.6	.003	15.4	.048
F	42.7	.111	63.8	.094	17.5	.151	59.0	.057	31.0	.007	18.5	.135
G	30.8	.093	55.0	.067	22.2	.069	49.3	.032	51.5	.007	12.3	.065
H	60.5	.090	137.0	.036	29.2	.155	146.0	.070	71.0	.010	26.8	.062

a = average number of items completed by group in given time of test. (See page 25.)
%e = percentage of errors to amount done—average for group.

TABLE IIIC

STANDARD DEVIATIONS FOR AMOUNTS DONE AND ERRORS

Grp	1	e	2	e	3	e	4	e	5	e	6	e
A	19.7	5.26	38.0	4.31	9.44	2.40	37.0	3.23	18.4	1.58	5.77	1.26
B	17.9	4.81	42.0	4.69	11.8	1.53	32.9	1.92	18.4	2.36	5.62	1.08
C	17.8	1.64	52.3	1.88	10.8	1.91	26.3	2.37	12.9	.47	5.31	.73
D	17.2	4.95	50.9	6.19	11.3	1.92	30.1	3.52	13.9	.82	6.65	1.26
E	11.0	2.66	41.8	2.42	8.61	2.49	22.9	3.03	13.5	.40	6.85	.77
F	19.0	3.58	32.6	4.34	10.1	3.41	23.3	3.52	12.0	.67	8.70	2.39
G	12.1	2.13	22.6	4.66	8.5	1.41	20.9	1.67	17.7	.60	4.16	.91
Co.[1]	19.4	4.43	46.8	4.60	12.2	2.58	37.8	3.83	19.5	1.47	7.23	1.51

[1] Co. = combination. See ¶ 2, page 57.

See pages 26 and 27 for information on make-up of groups A, B, C, D, etc.

The number at the top of the column refers to the "amount done" variable of the test of that number, while the following "e" refers to the errors for that same test.

How then should a selection be made? Which of the six tests of this study should be selected? Let this now be taken up as a problem. There is presented here no criteria for a group of so-called general clerical workers, each group represented having been composed of those performing more or less specialized duties. There might be selected, however, from among the operations performed by the clerks of these various groups, those that would, taken together, constitute what might be called general clerical work. In other words, the elements of so-called general clerical work are present in the work of these clerks. The problem in civil service would in cases be analogous to that of selecting prospective clerks to be assigned, in the order of their place on the list, to any one of these various types of work as the openings might occur. Then, suppose this to be the problem: The selection of clerks on the basis of their fitness to be placed in any of these various lines of clerical work. What would be the procedure?

Three factors stand out as important in determining which tests would best make this selection. The relative effectiveness of the tests for this purpose would tend to be affected by—

1. The frequency with which the fundamental operation, serving as the basis of a test, is found as a factor in the everyday work of the various groups of clerks.
2. The average degree of correlation between a test and the various criteria representing different types of work.
3. The degree of variability in the correlation of a test with the various criteria.
4. The degree of correlation with the other tests selected.

In summary form: That test would seem to be the best for a battery (to be used in selecting for certification to any of the types of clerical work) which had as its basis a fundamental operation most frequently found in the various types of work,—the highest average correlation with all of the criteria; the least degree of variability in its correlation with the different criteria; and the lowest intercorrelation with the other tests.

On the basis of these principles the six tests of this study may be considered. The relevant data are presented in Table IV.

The data in this table on page 62 are placed together for the sake of convenient comparison. A consideration of these data would seem to indicate that—

a. Test 6 is best on the basis of average high correlation.

b. Test 3 is best on the basis of small variability.

c. Test 4 is the poorest of the battery, all round.

d. Test 4 "amount" alone considered, is probably the best index of general intelligence, test 2 coming next.

e. The accuracy scores for test 2 give the highest single score correlation with general intelligence.

f. Tests 1 and 3 for amount correlate to the lowest degree with intelligence.

g. On the basis of low average intercorrelation with the other tests, test 3 takes the lead, followed by test 2 which shows the lowest correlation with test 3 of any correlation in the table.

TABLE IV

COMPARATIVE DATA FROM THE SIX TESTS
(GROUPS A–G INCLUSIVE)

Average Correlations; Ranges; and Average Deviations.

Correlations Between Aptitude and Test Scores with Best Weight for Errors					Correlations Between Aptitude and Test Scores (Amount Only)			
Test	Av. R_I	Range of R_I Min.	Max.	Variability	Av. R_I	Range of R_I Min.	Max.	Variability
1	.41	.06 to	.65	.16	.32	—.04 to	.51	.14
2	.44	.07	.74	.21	.31	.00	.72	.19
3	.39	.30	.53	.07	.36	.30	.52	.05
4	.33	.08	.52	.12	.20	—.20	.50	.19
5	.39	.25	.56	.10	.37	.12	.56	.12
6	.46	.17	.65	.16	.38	.10	.65	.18

Variability is the average deviation of the correlations from their average.

CORRELATIONS BETWEEN TEST SCORES AND INTELLIGENCE
(41 HIGH SCHOOL PUPILS)

Test	1	2	3	4	5	6
Amount Done	.15	.41	.23	.51	.27	.33
Accuracy	.24	.54	—.01	.16	.15	.18

Intelligence Scores taken from Army Alpha Test.

INTERCORRELATIONS BETWEEN "AMOUNT DONE" SCORES FOR ALL GROUPS—ALL SCORES FOR GROUPS A–H COMBINED

Tests	1	2	3	4	5	6
1		.40	.31	.38	.50	.42
2	.40		.24	.41	.41	.45
3	.31	.24		.32	.52	.26
4	.38	.41	.32		.48	.53
5	.50	.41	.52	.48		.40
6	.42	.45	.26	.53	.40	
Average	.40	.38	.33	.42	.46	.41

The correlations in the table were taken from those worked out from the combined scores of the eight groups. See Table III, also footnote of Table III.

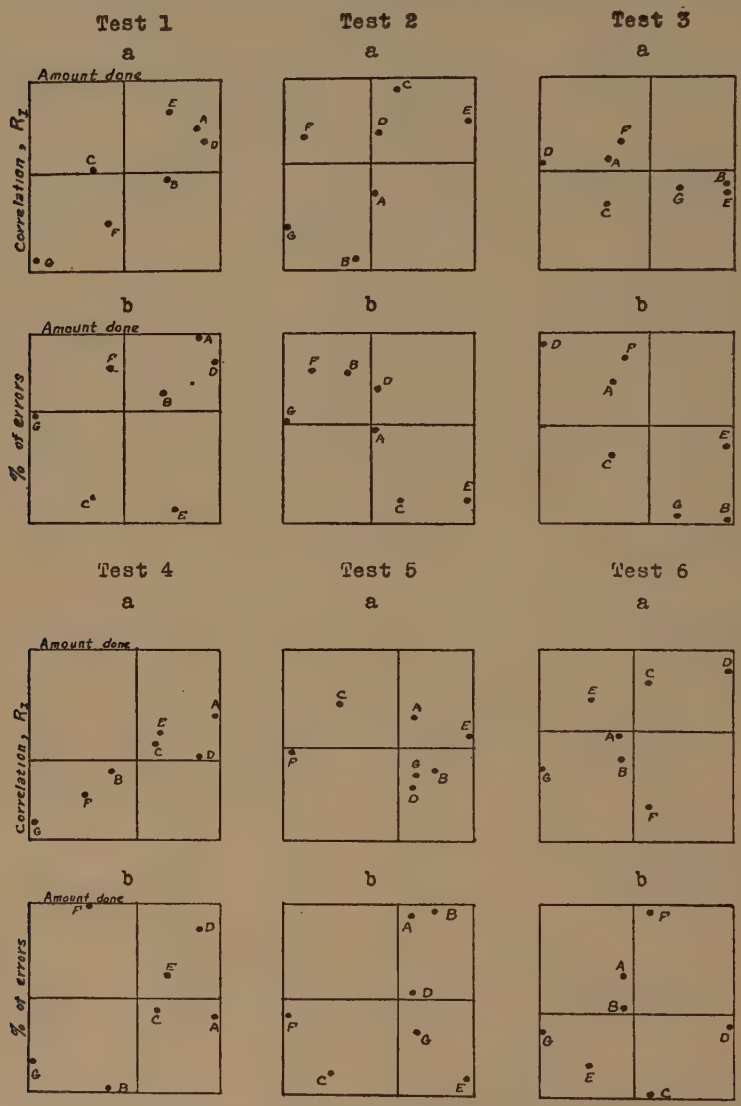
These are suggestive, but it must be remembered that the final selection depends upon a combining of the factors, suggested as useful, in the determination of the effectiveness of the various tests, for a battery for general clerical work. These data should, however, prove suggestive in the selection of any of these tests for use in a team of tests contemplated or already in process of construction. The data presented give the information necessary to judge of the value of any one test, as it might be selected, to add to another battery as suggested. A test with particular characteristics might thus be selected.

Such data as those included in Table III (page 58) and as shown in Table IV are capable of an almost endless study, and much that is being suggested here in a rather crude way, for the sake of emphasizing certain underlying considerations, might be treated statistically, did time permit. Many interesting suggestions for further study have rather naturally been suggested at many points. From the data given here, any interested reader may make for himself various other comparisons, and perhaps come to important and helpful conclusions. The data on test 3 may be taken as illustrative.

The above discussion showed the following concerning test 3: (1) It has the least variability of any of the tests; (2) its intercorrelation average is less than that of any other test; (3) the conclusion reached is that these would seem to indicate its value in a battery for testing ability to perform general clerical duties. Further suggestions concerning this test are found: (1) from Chart 10 on page 64, it is seen that for this test there is a tendency for the correlation with the criterion to be higher with those groups, the

CHART 10

DISTRIBUTION OF GROUP RESULTS
(SHOWING RELATIONSHIP BETWEEN)
A, MAXIMUM r AND AMOUNT DONE; B, AMOUNT DONE AND ERRORS (%)



The above dots are located on the basis of group averages.
The letters beside the dots indicate the groups.

amount of whose performance on the test is the smaller; (2) that there seems to be quite a distinct negative relationship between amount done and errors made—the smaller the amount done, the greater the percentage of errors. Then from Table IV it is seen that test 3 has next to the lowest correlation with intelligence, and that accuracy in the test has a practically zero correlation with intelligence (—.01). Such data as these lead to interesting thought and guesses, and may have in them valuable suggestions for further test construction.

CONSIDERATION OF TESTS FOR PROGNOSIS OF SPECIALIZED CLERICAL WORK

According to the same principle as that used in determining the value of a test, as fit for a battery for general clerical prognosis, it may be assumed that those tests showing the widest variations are likely to be the ones whose values vary according to the type of clerical work in question. As the tests were devised on the basis of samplings translated into what were called fundamental clerical operations, it would be expected that their correlations with the criteria of the different groups should show higher correlations in the cases of those tests that are based on the fundamental operations which, by analysis, are seen to be included in the actual duties of the clerks of such a group. An investigation of this point may now be made. In Table V are presented certain significant correlations for each group, and there is also given a brief analysis of the work of such groups. Each duty seeming to correspond with a given test is indicated opposite that test, and connected by a dotted line with the test—the test that represents the fundamental operation with which this duty corresponds. A brief study of this table will now be made.

Two factors enter to lessen the feasibility of treating the data of the table in a mathematical fashion with any assurance. The number of groups, including the same fundamental operation as that on which the test is based, is not large enough to make possible a satisfactory averaging. Second, there is a lack of the detailed knowledge of some of the duties which the writer desired, but which would have required a careful analysis, at first hand, of the detailed work of these offices, an analysis that was not possible under the conditions of this study.

In the Appendix is given a more detailed statement of the work

TABLE V

TYPES OF WORK PERFORMED BY GROUPS, AND TEST RESULTS

Group, Fundamental Operations	Test	R_I	r_{Ia}	r_{Ie}	$R_{I'}$	$r_{I'a}$	$r_{I'e}$	Averages Amt.	Averages % Errors
A Compares names and addresses.	1	.58	.34	—.13	.22	%.22	.11	57.3	.13
	2	.34	.20	—.24	.24	.05	—.23	80.7	.06
	3	.45	.34	—.20	.09	—.01	—.09	16.6	.14
	4	.52	.50	.19	.34	.31	.05	82.5	.04
Copies numbers	5	.51	.50	.08	.38	.11	.36	50.6	.01
Makes simple computations	6	.45	.44	—.17	.26	.26	—.03	16.8	.10
B Compares names and addresses.	1	.41	.40	.21	.01	—.01	—.01	52.0	.10
	2	.07	.05	—.02	.18	.06	.17	74.0	.09
Handles bills alphabetically	3	.35	.32	—.02	.05	.02	.05	27.0	.06
	4	.30	.24	—.08	.12	—.10	—.09	62.7	.03
Copies names and prices	5	.31	.30	.05	.28	.06	.27	54.3	.01
Makes simple computations	6	.34	.25	.30	.08	.04	.08	17.0	.08
C	1	.42	.38	—.09	.34	.16	—.25	40.2	.06
	2	.74	.72	.19	.52	.50	.09	92.9	.03
	3	.30	.30	.12	.13	—.09	.04	16.6	.10
Copies numbers as read	4	.41	.35	.34	.36	—.13	.25	70.2	.04
Looks up rates in tables.....?	5	.56	.56	—.06	.56	.55	—.18	38.0	.00
Makes simple computations	6	.65	.65	—.14	.57	.57	—.04	18.2	.03
D Compares Numbers	1	.56	.51	.11	.45	.28	—.12	59.0	.11
Counts	2	.58	.38	—.18	.64	.16	—.45	83.4	.08
	3	.44	.36	.00	.35	.21	—.10	11.2	.16
Sorts tickets	4	.36	.22	—.13	.20	—.01	—.18	77.0	.05
Makes tabulations	5	.25	.12	—.19	.22	—.06	—.22	50.8	.01
Makes simple computations	6	.70	.48	—.56	.64	.41	—.53	22.7	.07
E Checks salesmen's prices.....?	1	.65	.47	—.22	.42	—.07	—.41	52.8	.06
Issues manufacturing orders	2	.62	.47	—.23	.23	.21	.17	113.1	.03
	3	.34	.33	—.17	.11	.11	.00	27.0	.10
	4	.44	.37	.31	.31	—.03	.30	71.9	.05
Copies orders	5	.46	.46	.16	.15	.14	.11	59.6	.00
Makes computations	6	.59	.56	.03	.52	.28	.34	15.4	.05
F	1	.22	.17	.05	.24	.21	.09	42.7	.11
	2	.57	.33	—.45	.67	.43	—.50	63.8	.09
	3	.53	.52	.09	.58	.58	.19	17.5	.15
Looks up prices in table.....?	4	.20	—.20	—.15	.15	—.02	—.13	59.0	.06
Copies prices on cards	5	.37	.35	—.29	.54	.47	—.47	31.0	.01
Makes simple computations	6	.17	.17	—.04	.36	.36	.02	18.5	.14
G	1	.06	—.04	—.06	.21	.21	.10	30.8	.09
	2	.18	.00	—.18	.22	.21	.05	55.0	.07
	3	.33	.33	—.02	.17	.11	.14	22.2	.07
	4	.08	—.06	.01	.14	.12	—.01	49.3	.03
Copies from slips	5	.29	.29	—.15	.39	.39	—.15	51.5	.01
	6	.32	.10	—.26	.36	.05	—.36	12.3	.07

R = Multiple Correlation. I = Aptitude Criterion. I' = Efficiency Criterion. a = Amount Variable. e = Error Variable.

of the different groups, and for those who are familiar with certain of these types of operations, the table on page 66 should be of significance. Taking up one test at a time it will be possible, however, to get a general line on the suggested value of each test as it applies to different types of work, or as it is similar to the various fundamental clerical operations set forth at the beginning of this study. The analytical weakness of the table is not in what is stated, but rather in the probable absence of factors or operations among those set forth. And these factors may be quite dependent for their success upon one or more of the fundamental operations. For instance, the work of the Order Clerks is of a rather high grade. In this work, such a step as "making out an order" is no doubt rather complicated, and involves not any *one* fundamental operation, to such an extent at least as to make it apparent to one not in close touch with the work. And yet back of this work certain mental functions may be called into play that are the real determining factors in certain of these fundamental operations. This lack is due, as suggested, to inability to analyze carefully the job operations at first hand.

With this preliminary suggestion the tests will now be discussed in order, on the basis of the information found in the table.

Test 1

The checking operation involved in this test appears in a simple form in the duties of three of the groups, A, B, and D; in the duties of group E in a more complicated form, while in the duties of three groups it does not seem to appear at all as a part of the regular procedure. Did we have data on, say, twenty-five different groups, it would then be suggestive to average the correlations in the case of the groups whose duties included the checking operation, also those where it did not seem to enter, and compare the results. This method is followed in connection with this test, as indicative of the procedure with a larger number of groups.

The results show the following: Groups whose duties include the fundamental checking operation have correlations with the aptitude criteria of .58, .41, and .56, with an average of .52, while those not showing the presence of such operation have correlations of .42, .22, and .06, with an average of .23. With group E, where this operation enters, but in a less routine fashion, the correlation is .65, and has not been considered in either of the above divisions.

The results would seem to indicate that test 1 is a good selective test for clerical work involving checking of this nature, and offers a rather high average correlation.

Test 2

From the table it will be seen that in a direct, apparent fashion, the fundamental operation of comparing does not enter into the duties of any one of the groups, and yet this test shows a rather high correlation with most of the groups, with the exception of groups B and G. It is further noticed that with the groups where a more intricate checking or selective operation is involved, as in looking up prices or dividends in an elaborate table, this test shows an especially good correlation. On the other hand, with the two groups where the lowest correlation is shown, the work is of the simplest and most routine nature—no such discriminative selection or checking is required. It would seem then that this test may have an especially important value in connection with the less routine and less simple clerical work. The intercorrelation table, page 63, also shows that this test (when considered for amount only) has the next to the lowest intercorrelation average of any of the tests. Table IV shows it to have next to the highest correlation with general intelligence, while accuracy in the test has the highest correlation with intelligence of any of the test scores. In this connection it will be seen from the data of Table III that for six of the groups the errors of this test show a negative correlation with the criterion. These facts taken together would seem to indicate that this test is a test best suited for selecting clerks for the less routine positions.

Test 3

This test has already been given detailed consideration, and it will be necessary here only to suggest again the possibilities of its being a very desirable test for a battery of tests for selecting general clerical workers, or even in the selection of special workers if the investigator is satisfied with its comparatively low but constant correlation. None of the groups were engaged in this work of filing. With a group of 21 filing clerks in a state income tax office, doing routine filing, a correlation of .58 was obtained, as figured by the rank method, by the use of this test. The employes in this case were given only the one test.

Test 4

The operation of test 4 does not seem to enter directly into the work of any but one of the groups, and then as it is related to the sorting of tickets on the basis of "origin-destination." In this case it does not give a very high correlation, and is exceeded in the cases of two other groups where the fundamental operation of this test is not noted. The writer considers the underlying principle of this test good, but is dissatisfied with it in its present form. In its administration, quite a number of candidates did not read the direction carefully, and attempted to classify every item. In such cases, in scoring, no deductions were made for items classified which should not have been classified. This method was used consistently throughout. This, however, eliminated certain values of the test.

Test 5

The fundamental operation of this test enters into the work of every group, and the correlations are quite satisfactory, with not a very large variation. However, the correlations as they stand are hardly comparable, because of the numerous forms in which this operation may appear in the work of the different groups. Here also as in other tests the variation in the relative time occupied by the operation in such work, as well as the variation in the emphasis or significance of it in the work of the group, reduces its comparability. The test as it stands involves really three operations instead of one: First, finding the right item in the original table; second, transferring it correctly, i. e., reading it correctly, and remembering it during a brief interval; and lastly, locating it in the proper place in the tabular form. The test in this form has a real value, but in revising these tests and at the same time following the underlying principle of fundamental operations, about three tests would be substituted in place of this one.

Test 6

This operation appears in the work of all the groups but one. In most of the cases the correlation is satisfactory. This would seem, in its present form, to be a good test for most types of clerical work. Here again, a more detailed statement might be made, if the relative emphasis on this work and the relative time devoted to it were known.

SUGGESTIONS FOR IMPROVING THE TESTS

Inasmuch as clerical tests are now in the early stages of development, and as this study will only serve as a step in this development, and makes no pretense of offering anything in the guise of a final word, it seems in place at the close of the study to consider the possibilities of increasing the predictive value of such tests. In this connection several points will now be considered.

1. The possibility of improving the make-up of the tests themselves is apparent. In conformity with the method of approach, taken in this study, further refinements in the analysis of the work in question suggest themselves, such as would result in better samplings for the test, as to both content and form. There is also room for improvement in the content of and the method of giving directions. The directions which seemed to cause difficulty were those in connection with the alphabetizing test where in some cases, especially with one group made up of men (not included in the data presented), a number did absolutely nothing on the test, but spent the full time trying to make out what was to be done. It would seem that as the ease of understanding directions becomes less, the greater is the extent to which the elements of intelligence and experience enter as factors in determining the results of the test performances. And thus the results are less satisfactory, if it is desired solely to determine alertness and accuracy in performing a specific piece of clerical work. In later experiments with the tests, familiarity with the test material, and a comprehension of directions on the part of those tested, were provided for before the test proper was started.

2. It is the opinion of the writer that in the case of such tests as those of this study, a lengthening of them would be of real advantage. This would tend to make them approach more nearly to actual office conditions which, in sampling tests of this type, would seem advisable. Although not worked out for these tests it would be well, before proceeding with any revision, to determine the best length for each test on the basis of the data available. By knowing the reliability of a test as administered within a given time, it may be determined how many times as long the test should be in order to have a given reliability or to make possible a certain desired correlation between test and criterion, provided such correlation were possible through the use of the test under consideration.

3. Attenuation. Tests vary in their results from time to time, due to varying chance conditions. Such introduction of chance reduces the resulting correlation. The method used in determining the theoretical value of a test when this factor is eliminated is that of giving two forms of the same test to a group under as nearly the same conditions as possible. With the resulting data and the use of Spearman's formula for correction for attenuation, a theoretical correlation coefficient may be obtained, which is indicative of the possibilities of a test. It is the correlation possible with a true criterion and a true test score. This method was not used in connection with the present study, but is suggested to those who have the data and the opportunity to make use of this method of correcting for attenuation. As factors in the determination of this correction for attenuation, it is necessary to have the reliability coefficient, secured from the data resulting from the giving of the two forms of the test in question. This involves the problem of the construction of two such similar tests. In this connection the writer again draws from the very helpful suggestions made by Kelley. Discussing the comparability of two tests he sets up the following rules for guidance in the construction of such tests:

1. Sufficient fore-exercise should be provided to establish an attitude or set, thus lessening the likelihood of the second test being different from the first, due to a new level of familiarity with the mechanical features, etc.
2. The elements of the first should be as similar in difficulty and type to those in the second, pair by pair, as possible; but
3. should not be so identical in word or form as to commonly lead to a memory transfer or correlation between errors.

4. The administration of the test is also capable of improvement. Those to whom it is given should be able to work under favorable conditions as to desk space, etc. Small wooden desk pads, such as are used in school class-rooms must sometimes be used, but are not so satisfactory as a broad desk or table where clerks can arrange things as would be done in actual office work. The extent to which variations in such conditions affect results is of course not known. Although it might not affect to any considerable extent, the validity of a correlation for an individual group, it might introduce an element of error in the comparison of results of different groups. The time of day was not the same for the different groups and would, no doubt, be quite difficult to arrange, but this also

CHART 11

TEST 1: DISTRIBUTION OF SCORES FOR 1000 CASES

Amount Done

Errors	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Row Totals	Cumulative
19																12	11	10			33	1,000
18														1	1	4	15	4			25	967
17																1	4	1			6	942
16													1	1		6	5	2			15	936
15																8	1	7			13	921
14																		6			6	908
13												1	1	2	1	4	6	1	1		17	902
12												1		2	1	1	2	2	3		12	885
11											1		2	3		3	4	2			19	873
10										1	1	1			3	5	3	4	1		19	854
9										2	1		3	5	7	3	2	6	3		32	835
8										1	2	3	11	0	6	2	2	5	4		37	803
7							1	1	1	2	5	4	2	5	7	5	3	3	2		41	766
6					1		1	1	1	7	11	3	5	6	11		1	3	5		56	725
5						1	5		2	8	17	7	5	11	11	1	4	2	3		77	669
4						1	4	3	7	14	29	10	8	11	7	7	6	5	1		113	592
3				1	1	3	3	4	5	27	21	18	12	10	5	5	2	2			119	479
2	1		3	4	2	8	4	13	35	25	16	16	9	8	2	6	3	1			156	360
1		1		8	3	17	6	11	24	16	16	5	8	1	3						129	204
0	7		2	11	9	24	3	3	6	5	4			1							75	75
Column Totals	8	1	6	25	20	63	22	44	130	133	81	73	85	71	45	65	87	41				
Cumulative	8	9	15	40	60	123	145	189	319	452	533	606	691	762	807	872	959	1000				

CHART 12

TEST 2: DISTRIBUTION OF SCORES FOR 1000 CASES

Amount Done

Errors	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	Row Totals	Cum. Row Totals
19					7	5	3	3	5	3	1	5	1	4	2	2	2		5		48	1,000
18					1			2											1		4	952
17			1		1			1			1				1	1					6	948
16				2		1	2	1	2	1	1								1		11	942
15		1			3	1		2	1					1							9	931
14			1	1	2	1	1	3			1	1									11	922
13		2		2		3	2	1	1		1	1		1							14	911
12			1		1	2		2	1		1		1								9	897
11		5	1	2	1	1	1	2	2	1		2	1				1				20	888
10				3	5			1	1	2			3						1		16	868
9		2	1	4	2	2					2		1	1							15	852
8		1		2	1	3	1	3	3	5	3		1				4				27	837
7	2	2	2	6	2	1	3	2	3	4	3	1	1						3		35	810
6	1	2	2	4	7	3	5	4	2	6	1	3					1				41	775
5	2	1	3	9	6	7	2	5	1	8	2	4	1		1				1		53	734
4	1	3	5	7	16	13	9	6	10	6	2	5	4	2	1	2			2		94	681
3	1	3	2	11	14	20	9	6	7	7	9	9	2	1			1		3		105	587
2	1	5	6	16	13	24	16	9	10	7	10	6	6	6		2	2	1	2		141	482
1	2	3	8	28	23	29	25	17	13	4	14	9	3			2	2		1		183	341
0	19	7	15	20	22	18	21	11	10	3	6	4		2							158	158
Column Totals	24	26	53	95	128	142	109	68	85	46	69	53	27	24	7	10	13	1	20			
Cum. Totals	24	50	103	198	326	468	577	645	730	776	845	898	925	949	956	966	979	980	1000			

CHART 13

TEST 3: DISTRIBUTION OF SCORES FOR 1000 CASES

Errors	0	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	Row Totals	Cumulative
19																						
18																						
17																						
16																						
15																						
14																						
13																						
12																						
11																						
10					1	2															3	1.000
9									1	1											2	997
8																					0	995
7			3	1			1			1											6	995
6			5	2	2		1	1	1	1											13	989
5		2	5	2	2	4	3	1	1	1	2										23	976
4		7	6	4	4	4	6	4	2	5	1		1								44	953
3		5	6	16	11	7	13	2	3	3	2	1	1								70	909
2		41	31	82	22	32	32	41	21	10	3	5	2	1	1	2					163	839
1		82	82	63	93	43	24	21	11	7	4	3		1							238	676
0	1	45	24	44	48	56	40	31	23	10	6		3		4		2				438	438
Column Totals	157	79	113	134	135	112	103	64	39	28	14	9	3	6	0	2	2					
Cumulative	157	236	349	483	618	730	833	897	936	964	978	987	990	996	996	998	1000					

CHART 14

TEST 4: DISTRIBUTION OF SCORES FOR 1000 CASES

Amount Done

Errors	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	Row Totals	Cumulative
19				1	1	1	1			2		2	3	1	2		1	1	1	5	32	1.000
18									1						1						2	968
17											1	2									3	966
16									1												1	963
15											2		2								4	962
14					1	1				1	1		1		1	1					7	958
13									1	1											2	951
12							1		1		1			2							6	949
11				1					3	1	1	2									8	943
10								2	3	3	1	2	1		1				1		14	935
9						1		1	3	2		2			1						10	921
8							3		5	2	2	3		1		1			1		18	911
7						1	4	5	5	4	3	1				1	1		2		27	893
6					2	3	2	7	8	10	3	2	2	3		1			1		44	866
5					2	5	8	6	12	4	5	2	3	2		2	2				53	822
4			1	1	6	8	10	10	13	15	5	13	4	4	5	1	1				97	769
3		1			2	9	17	13	11	18	18	9	8	3	6	2	2	2		2	123	672
2		1		3	8	19	18	17	19	24	17	14	11	7	7	2	1	2			170	549
1		1	1	1	8	15	21	24	23	22	17	11	15	6	4	7	2	1	1	1	180	379
0		12	4	9	18	22	41	23	21	15	9	8	6	4	4	3					199	199
Column Totals.	15	6	22	52	85	121	104	107	132	90	78	55	32	41	14	13	9	1	23			
Cumulative	15	21	43	95	180	301	405	512	644	734	812	867	899	940	954	967	976	977	1000			

CHART 15

Test 5: Distribution of Scores for 1000 Cases

Amount Done

Errors	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	Row Totals	Cumulative
19													1		2						3	1,000
18							1														1	997
17																					0	996
16					1		2				1	1									5	996
15													1								1	991
14																					0	990
13															1						1	990
12									1												1	989
11																					0	988
10																					0	988
9		1							1	1											3	988
8																					0	985
7				1				1	1												3	985
6			1	1				1		1	1		1								6	982
5			1					1	1						1						4	976
4	1		1		1	1	1	1	1		1			2	1						11	972
3		1				3	2	3	1	2	5		1	2						1	21	961
2			1		1	7	6	6	7	7	9	5	2	5	1	1	1				59	940
1		1	3	6	11	6	2	13	12	9	3	1	2	5	1	4	5	1	2		207	881
0	16	4	4	14	36	36	49	90	88	68	69	70	40	51	32	15	15	7		3	2	674	674

Column Totals	17	7	11	23	49	56	123	147	108	111	111	59	72	46	24	19	9	2	3	3	
Cumulative	17	24	35	58	107	163	286	433	541	652	763	822	894	940	964	983	992	994	997	1000	

CHART 16

Test 6: Distribution of Scores for 1000 cases

Amount Done

Errors	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	Row Totals	Cumulative
19																					0	1,000
18																					0	1,000
17																					0	1,000
16																					0	1,000
15													1								1	1,000
14																					0	999
13													1								1	999
12											1										1	998
11																					0	997
10																					0	997
9																					0	997
8										1						1					2	997
7													1								1	995
6						1				1	2	1		1	1	1			1		9	994
5					1	1	1	1	2	1	3	1	2	1	1	1	1			1	18	985
4					1	1	4	2	3	3	4	5	5	2	4	3	1		2	1	41	967
3					1	3	2	6	15	6	11	4	7	2	3	2	1	4	2	2	1	73	926
2					7	7	10	16	30	15	24	20	22	8	7	3	3	4	1	2	2	181	854
1			2	3	13	16	15	23	48	18	49	24	22	11	16	12	11	3	2		1	291	673
0	13	1	6	19	16	21	26	50	44	60	28	40	14	10	11	7	6	5		4		382	382

Column Totals	13	3	9	42	47	53	73	148	87	153	85	100	40	44	33	26	17	10	7	10	
Cumulative	13	16	25	67	114	167	240	388	475	628	713	813	853	897	930	956	973	983	990	1000	

in the comparison of results of different groups may introduce a factor influencing reliability.

STANDARDS FOR JUDGING TEST PERFORMANCE

Inasmuch as this study has emphasized the opinion that it may be best to give different weightings to the same test, when used as a selective instrument, for different groups, no attempt has been made to standardize the tests on the basis of a common weighting. Further, considering the present status of clerical testing, it has been thought best to present all data, in so far as possible, in such shape that they might be applicable in the solution of individual office problems. To this end the distributions of the 1,000 persons used in the standardization were made on the basis of both amount done, and errors. These distributions appear in Charts 11 to 16. In the body of each table is shown the number of persons performing a given amount with varying numbers of errors. The scale for errors appears at the left while the amount done appears at the top. The class intervals used here are the same as those established for transmuting the raw scores of the data, for the various groups, into steps for determining the correlations. In connection with both the errors and the amounts done, the totals of the various arrays are given as column and row totals, but a scale showing the cumulative score is also given in each case. This last is shown for the benefit of any who, in using the tests, may like to determine the percentile score of an individual in terms of the distribution of the 1,000 cases represented in the distribution tables. Thus for illustration, in Chart 12 for test 2, a person covering 53 items with 1 mistake would be in the group of 142 persons accomplishing an amount between 50 and 59, there being 32.6 per cent doing less than that amount, as shown by the cumulative totals. But as to errors, it is seen that such person would be superior, as those making one error or less constitute only 34.1 per cent of the total group of a thousand. Thus such person is below the average in amount done and also below the average in errors made, which is really above the average in accuracy. If desired, these scores may be turned into percentages of errors made, as was done for the small groups for whom criteria were secured. These tables are presented only as suggestive to students of the subject, and not as final, as the writer believes that certain changes should be made in the test before a final standardization is worked out.

NOTE: The data used in making this distribution chart were taken from about 800 scores of candidates in a civil service examination in New York State for "Junior Clerk" and for "Clerk"; from scores made by employes in the service of the state of Wisconsin; and from part of the scores made by clerks in the groups considered in this study. The make-up of the group of 1,000 is therefore heterogeneous, made up of both experienced and inexperienced persons. It is given only as suggestive of the trend of distribution. In the New York State examination, for tests 2 and 4, a penalty of 5 errors was given when candidates did not follow directions absolutely, while in the other groups no deduction was made for these violations of directions, as when a person attempted to classify all items in test 4. (See page 69 for discussion). This would, then, in the distribution for errors in these two tests, tend to raise the average slightly above what it would have been, had all been marked on the same basis—that used by the writer.

CHAPTER VI

SUMMARY

In the nature of a partial summary and closing suggestions pointing to further study, the writer desires, in conclusion, to leave the following thoughts with the reader.

1. A comparison of the test correlations and the duties of the clerks who were tested seems to indicate a higher correlation between criteria and tests, where both work and test involve the same fundamental operation. This statement cannot be backed by the definite data that might be desired because of the impossibility of studying the work of the clerks at first hand as was done with the clerical work of the state service. With the point of view and suggestion given in Chapter I, an office manager having under his supervision a large group of clerks would be in a position to make such a thorough analysis and adequately compare the results with the showing made by the clerks on the various tests. Through the accumulation of such data, progress in clerical test development will be greatly facilitated.

2. In trying out these or similar tests in a given office, not only should a thorough analysis be made of the work of the clerks but very carefully considered ratings should be made up. With such data the results of the tests may be worked out according to the methods suggested and used in this study. In this way the best weightings for errors and the best test combinations may be determined. The advisability of adopting the same weights or test selection determined for the groups of this study is questionable, even though the jobs carry the same titles as, for instance, "pay-roll clerk." The operations actually performed by the pay-roll clerks vary from office to office. A study of the analyses in the Appendix will make this evident.

3. The assignment of the best weighting for errors is shown to be a real and important problem. The curves presented in this study in Charts 6 to 8 show marked differences in correlations resulting from comparatively slight changes in weighting for errors. As has been pointed out on page 51 and as indicated in the equation used (page 45), the weighting and the suggestive curves are depend-

ent upon the zero order correlation coefficients entering into their make-up. The curves in Chart 9 show large variations for groups supposedly the same, due to large variations in the zero order coefficients. This then makes it questionable whether a weight worked out for one group will be best for other similar groups. This is not a very satisfactory conclusion to draw, especially when it is considered that in every test given, errors *must* be weighted. The method suggested is probably the best to use, however, (with caution) until further light has been thrown on the matter by the statistician.

4. The writer believes that progress in the development and practical use of clerical tests will best be served by giving the greatest attention and emphasis in study to the individual test and to the types of work for which it shows the highest selective value. The combinations or batteries may later be made up to meet the needs of different offices. With this in mind, it would seem well to lengthen the individual tests and organize them as units, each test in a separate folder.

5. In connection with all groups but one, two sets of correlation coefficients are given: one for the "aptitude" criteria and the other for the "efficiency" criteria. The idea of the writer in constructing the tests of this study was not that they might function as trade tests—tests of present efficiency in the work of a job, immediately applicable—but rather that they might function as tests of aptitude and be of value in selecting clerks from among those without any actual experience in clerical office work. At the conclusion of the study, the writer believes them to be of value as tests of aptitude, although the present data do not permit of making such a statement authoritatively. The data are, however, very suggestive.

In thirty of the forty-two pairs of correlations with the two criteria for seven groups, thirty pairs show a higher correlation with the aptitude criterion than with the efficiency criterion and one pair the same correlation. The exceptions are concentrated in two groups, for of those not showing this relationship, eight pairs are found among two groups, groups F and G—a pricing clerk group and a pay-roll clerk group.

For all four groups for which the writer had data on length of employment, the second or aptitude ratings when used as criteria showed only about half as high a correlation with length of employ-

ment for each group as did the ratings on efficiency when thus used. (Table I, page 29). The facts appear to indicate that the aptitude ratings come nearer to representing the criteria which the writer wished to secure than do the efficiency ratings. If so, then the tests themselves would seem to be more effective in determining aptitude or possibilities in clerical work than in determining clerical achievement for immediate functioning. The differences in the correlations are shown in Table 5, page 66.

Further study needs to be made along this line—study which the present data do not make possible. It is suggested that this matter might be quite largely cleared up by giving the tests to inexperienced applicants for positions in a certain office and then following through their records for several years, repeating the test in other series from time to time with the entire group of employes in question, and also providing opportunity for trying out those who seemed to make poor showing in the tests as first administered.

In this connection it should be noted that certain factors enter to complicate the problem. What is really wanted in the employes selected? Is it for the best interests of the office to select those who have possibilities of growth or those who will immediately enter into the work with the highest degree of efficiency, irrespective of possible future growth in the work? This must of course be decided upon for the individual office and is dependent to a large extent upon the average tenure of service in the job as well as upon the organization and promotion plans of the concern in question.

Further, what is to be done when, as is so frequent in civil service examination, groups of applicants are quite heterogeneous in their make-up, some persons having had considerable office experience and others having had none? These are practical problems and must be studied in a serious manner with proper control conditions. It was not possible to include them within the scope of this study but such a group of tests as the one here presented, ought to serve well as an instrument in the study of these problems.

6. In Chapter V are given the conclusions reached relative to the values of the different tests, and a summary here would hardly add to the suggestions made in the detailed discussion which those interested in using the principles in any of the tests will naturally go over carefully, as they also will the various tables, and the job analyses found in the Appendix. It may be well, however, to point out here that the tests seem to have a wider range of applicability

as selective instruments than the writer had in mind at the beginning of this study. In this respect, however, the individual tests seem to differ. Test 2, for example, shows an especial value in determining aptitude for the more complicated clerical duties, as has already been suggested, thus reaching a higher level in its applicability than at first was under consideration. On the basis of the grading used in the clerical performance scale of Chapter I, these tests seem to have a selective value for the first two grades and to a certain extent for the third grade. This again indicates the wisdom of first considering such tests as these individually instead of as a predetermined battery.

7. In conclusion and in conformity with the suggestions made in the preceding paragraph, it should be pointed out that the greatest value in this study, especially for those working on clerical test construction, will probably be found in the tables presented, in which the various intercorrelations and criteria correlations of the zero order are given. With this in mind these tables are presented in the detailed form in which they are found.

APPENDIX

STATEMENT OF DUTIES OF EMPLOYES TAKING THE TESTS

GROUP A

The bill-ledger typists or clerks use the Elliott-Fisher bookkeeping and billing machine, in the work of preparing customers' light and power bills, making debit entries on the customers' ledger sheets, and making recapitulation sheets of these postings.

Aside from the manual operations are the following: The operator compares the name shown on the first ledger sheet with the name on the bill, and compares the meter number on the ledger sheet with that on the meter reading sheet. He then enters the present and previous months' readings, taking them from the meter book. The next entry is that for total consumption, which is automatically computed by the machine. The total consumption is then compared with the total consumption shown on the meter reading sheet. The operator then enters the consumption for the first thirty hours' use of the maximum, as taken from the meter reading sheet, then that for the second thirty hours, and finally the excess over the sixty hours.

The operator then enters the gross amount on the bill. This amount is obtained from the meter reading book, having been previously computed by the extension clerks. The discount is then entered. This is computed by adding mentally the consumption at the first and second rates. The net amount of the bill is automatically computed by the billing machine and is entered by the operator on the bill proper and then copied onto the stub of the bill.

GROUP B

The billing clerks, using Elliott-Fisher and similar machines, transfer from the charge and credit slips to the bills (in triple form including remittance slip) the items, date, price, and sales identification number, as well as the name and address of the customer. In adding items to a bill, the name and address of the customer, as shown on the charge or credit slip, must be compared with that on the bill as withdrawn for entries.

GROUP C

The dividend clerks assist in the general calculation of the dividends, computing and assigning to the various policies the appropriate dividends and equivalent values in paid-up insurance, and accumulating dividends at interest and in accordance with the contract provisions. The mathematical operations—multiplication, division, and addition are mostly mental. The work requires a degree of special training and technical knowledge of the dividends assignable to various contracts issued by the company since organization.

The dividend clerks *check* in pairs, one watching the entries while the other

calls back the results as found in the tables, and based on the data given by the first clerk. There are sixty-four schedules used in this work, of course with varying degrees of use. The checker is supposed to catch by observation any inconsistencies of figures.

GROUP D

This group is made up of those doing several types of work but with the larger number doing mostly the same type or similar work. *Interurban audit clerks* sort and count tickets and transfers, and balance the results against conductors' count. They also tabulate tickets as to point of origin and destination, calculating the value of tickets between particular stations, and audit conductors' receipts by calculating and tabulating the value of each receipt.

The express audit clerks sort way-bills numerically, with regard to point of origin and destination; tabulate way-bills on abstract sheets; and prepare monthly statement or summary by listing items from abstracts on adding machines, balancing adding machine totals against footup of abstracts and rechecking for missing way-bills.

The other work done by a few of this group includes calculating and tabulating the number of miles made by the cars on each line, and checking over the value of tickets turned in against the conductors' remittance sheets.

GROUP E

The order clerks are engaged in a high class of clerical work, involving a checking of prices shown on the orders sent in by the salesmen which requires the figuring of prices from a base with differential charges. A technical knowledge of construction of garments and character of material is necessary. For instance, certain fabric numbers can be made in one style construction only—a fancy plaid back overcoat fabric must not be sold full-lined.

These clerks also issue manufacturing orders where accuracy is of prime importance (no checking is done), and the economical cutting of materials depends upon the order combinations made and judgment shown. It demands absolute accuracy in copying with less of the selective and judging function. Accuracy in computing is necessary.

GROUP F

The pricing clerks assign piecework prices to an average of 15 operations. The work is highly technical, and involves an accurate knowledge of tailoring and construction of garments. New pricing clerks are sent into the factories for observation. The tables of piecework prices and differentials are very complex. For illustration: there are about 50 different prices and variations for the finishing of a sack coat; the proper price for that particular model and character of work must be correctly determined and written in the proper space on the piecework card. Speed is essential as the work is on a time schedule. The manufacturing instruction statements (one for each lot) are given to the pricer, who reads the description, and then prices the operation, using a price list. For buttonhole making the statement would show the number and size of buttonholes; the pricebook shows the price for that size of buttonhole, which is then

multiplied by the number of buttonholes indicated on the order, and the result entered on the coupon.

GROUP G

The payroll clerks compile the payrolls. This consists in copying in summary form on the payroll sheets, quantities of work done at various prices, from the piecework reports turned in by the factory. In addition to calling for absolute accuracy, variations in quantity of work and prices used must be watched for; that is, there are distinct relations in piecework prices in the group. For instance, if prices are running $3.50 and $4.50, a price of $10.00 would be distinctly wrong, or if 300 garments constituted a day's work, a record of 700 would be a mistake. Payrolls must be compiled on schedule time. The clerks perform no computations. Their work is straight copy work.

GROUP H

The payroll clerks make up the payrolls. The time-clock cards, showing the attendance registrations are received from the shop, and the total hours worked each day, and weekly totals, computed. Job tickets received from the shop are checked with the time cards to see that total hours agree. Daywork hours are posted from the time cards to the payroll. Also, daywork hours (or combination of daywork and piecework hours) are extended and totalled on each employe's account.

DIRECTIONS FOR GIVING TEST

The value of these tests depends largely upon the manner of giving them. Be sure to observe the time limits to the second, setting your watch so that the second hand agrees with the minute hand.

Before distributing the folders, make the statement that they are not to be opened until directions are given.

Distribute folders with "Directions" face up. Tell competitors to fill in the blanks at the top of the front sheet—allow about a minute for this.

Tell candidates to read directions carefully—allow two minutes for this.

At the end of two minutes, tell candidates to copy quotation on the lower part of the page as directed—allow two minutes for this.

After candidates have finished copying, make the following statement:

> "Remember that you are not expected to finish any of the exercises in the folder. When you are asked to begin on the next exercise, place a line under the last item that you have finished in the exercise on which you are working, turn the page and begin at once on the following exercise."

Allow for each exercise the time indicated below. There must be no lapse of time between tests.

Fill in the blanks below with the time of commencing on each exercise and return this sheet with your examination papers.

Exercise 1 ..... 5 minutes. Time commenced.....
Exercise 2 ..... 5 minutes. Time commenced.....
Exercise 3 ..... 8 minutes. Time commenced.....
Exercise 4 ..... 5 minutes. Time commenced.....

Exercise 5 ..... 7 minutes. Time commenced.....
Exercise 6 ..... 3 minutes. Time commenced.....

When it is time to stop one exercise and begin another, make the following statement in a quiet way:

"You will now begin the next exercise."

After the competitors have finished exercise number 6, make the following statement:

"You will now turn the folder over with the first or direction page face up."

Collect all folders immediately. NOTE: *Do not keep your watch in prominence. Give directions in a quiet manner. In timing the exercise be exact, but do not give candidates the idea of a race.*

DEFINITION OF TERMS USED IN THIS STUDY

1. In the field of vocational testing there are four forms of tests as set forth by Hollingworth (H. F.) in his book on *Vocational Psychology*, Chapter V, pp. 109–114. They are as follows:

1. *Vocational Miniature:* "Here the entire work or some selected and important part of it is reproduced on a small scale by using toy apparatus or in some way duplicating the actual situation which the worker faces when engaged at his task." A miniature switchboard for testing telephone operators is used as an illustration.

2. *Sampling Tests:* "Closely related to this method of miniature performance is that of taking an actual piece of the work to be performed and sampling the candidate's ability by his success in this trial. Thus, in connection with the recommendations of clerks and assistants from among the boys in commercial high schools it is common to test their ability from time to time throughout their course by assigning them small pieces of work similar to that which they might later be required to perform in business offices and shops. Finding addresses and numbers in a telephone directory, carrying out involved directions from memory, computing calculations, recommending action on the basis of their figures, making out a trial balance, a trial chemical analysis, etc., are common forms of this type of test. In certain cases such specimens of work have been devised in or taken into the psychological laboratory and the worker watched more closely and measured more exactly. This has been done for instance by Thorndike in the case of clerical workers and salesmen, by Paynter in the case of judges of trade-mark infringements, by Scott in the case of salesmen, and by others in the case of tests for handwriting experts."

The sampling test is the type of test used in this study and referred to on page 17.

3. *Analogy Test:* An analogy test is devised on the basis of its "real or supposed resemblance to the sort of situation met by the worker in the given occupational activity." A certain reaction test might be used because of the common "mental processes or psychological functions" involved in both test and the actual work in question.

4. *Empirically Selected Tests:* Tests of this type are usually selected from a larger number which have been tried out in an experimental fash-

ion. The tests may have no apparent suggestion of, or connection with, the work in question. They have been sought "by purely haphazard and empirical ways." Those are retained which show satisfactory correlations with the vocational ability under consideration.

2. *The analysis* or that part of the study described in Chapter 1, was made for the purpose of selecting specimens of clerical work to use in the make-up of the sampling tests. No attempt was made to present anything in the nature of job specifications. Such specifications would have included statements of working conditions, necessary mental and physical qualifications of the clerks, previous experience required, possibilities of promotion, salary ranges, etc. A complete job analysis of clerical work has been made by the writer, but in this study only that part of it is described that bears on the problem of devising sampling tests—the analysis of the actual duties of clerks—the objective clerical operations involved.

3. The term "*Clerical Performance Scale*" as used in this study might properly be replaced by the expression "Scale of Clerical Duties" or by "Standard List of Clerical Duties Arranged in a Scale According to Degree of Difficulty." The reference is therefore to illustrations or exhibits of actual pieces of clerical work such as might be performed by a clerk in connection with the regular duties of an office.

4. "*Fundamental Operations*" as used in the text of the study may be defined as simple, objective unit operations which through repetition and combination one with another make up clerical duties as they are ordinarily understood. They are the elements, the common factors, of clerical duties—the result of analyzing clerical duties into their simple objective operations or parts.

"Fundamental operations" and "basic operations" are used interchangeably.

Date Due

FEB 18 '53			
[illegible] 5 [illegible]			
JA 19 '55			
[illegible]			
[illegible]			
JUL 26 '67			

Remington Rand Inc. Cat. no. 1139.

HF5501
.R8

76878

O6-CPD-552

JAMES C. KIRKPATRICK LIBRARY - UCM
3 7250 04271913 6